

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 ; Search time 56.6732 Seconds  
(without alignments)  
1736.455 Million cell updates/sec

Title: US-09-941-947a-6

Perfect score: 3216  
Sequence: 1 MKLTDPYLNKHTPADIR.....LSLVGHDSKGLATIEQPCA 620

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDSI1/gcgdata/geneseq/emb1/AA1980.DAT.\*  
2: /SIDSI1/gcgdata/geneseq/emb1/AA1981.DAT.\*  
3: /SIDSI1/gcgdata/geneseq/emb1/AA1982.DAT.\*  
4: /SIDSI1/gcgdata/geneseq/emb1/AA1983.DAT.\*  
5: /SIDSI1/gcgdata/geneseq/emb1/AA1984.DAT.\*  
6: /SIDSI1/gcgdata/geneseq/emb1/AA1985.DAT.\*  
7: /SIDSI1/gcgdata/geneseq/emb1/AA1986.DAT.\*  
8: /SIDSI1/gcgdata/geneseq/emb1/AA1987.DAT.\*  
9: /SIDSI1/gcgdata/geneseq/emb1/AA1988.DAT.\*  
10: /SIDSI1/gcgdata/geneseq/emb1/AA1989.DAT.\*  
11: /SIDSI1/gcgdata/geneseq/emb1/AA1990.DAT.\*  
12: /SIDSI1/gcgdata/geneseq/emb1/AA1991.DAT.\*  
13: /SIDSI1/gcgdata/geneseq/emb1/AA1992.DAT.\*  
14: /SIDSI1/gcgdata/geneseq/emb1/AA1993.DAT.\*  
15: /SIDSI1/gcgdata/geneseq/emb1/AA1994.DAT.\*  
16: /SIDSI1/gcgdata/geneseq/emb1/AA1995.DAT.\*  
17: /SIDSI1/gcgdata/geneseq/emb1/AA1996.DAT.\*  
18: /SIDSI1/gcgdata/geneseq/emb1/AA1997.DAT.\*  
19: /SIDSI1/gcgdata/geneseq/emb1/AA1998.DAT.\*  
20: /SIDSI1/gcgdata/geneseq/emb1/AA1999.DAT.\*  
21: /SIDSI1/gcgdata/geneseq/emb1/AA2000.DAT.\*  
22: /SIDSI1/gcgdata/geneseq/emb1/AA2001.DAT.\*  
23: /SIDSI1/gcgdata/geneseq/emb1/AA2002.DAT.\*  
24: /SIDSI1/gcgdata/geneseq/emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1          | 3216   | 100.0       | 620    | 23    | ABG61581    |
| 2          | 3216   | 100.0       | 620    | 23    | AAE22301    |
| 3          | 3216   | 100.0       | 620    | 23    | AAU80325    |
| 4          | 2069.5 | 64.4        | 627    | 22    | AAU00012    |
| 5          | 2069.5 | 64.4        | 627    | 23    | AAO21846    |
| 6          | 2063.5 | 64.2        | 627    | 22    | AAU00014    |
| 7          | 2061.5 | 64.1        | 627    | 22    | AAU00013    |
| 8          | 2014.5 | 62.6        | 620    | 22    | AAU8450     |
| 9          | 2000.5 | 62.2        | 620    | 20    | AA52832     |

|    |        |      |     |    |          |
|----|--------|------|-----|----|----------|
| 10 | 2000.5 | 62.2 | 620 | 20 | AAV08880 |
| 11 | 2000.5 | 62.2 | 620 | 21 | AAE20544 |
| 12 | 2000.5 | 62.2 | 620 | 21 | AA51613  |
| 13 | 2000.5 | 62.2 | 620 | 22 | AAU34475 |
| 14 | 2000.5 | 62.2 | 620 | 23 | AAO21860 |
| 15 | 1937.5 | 60.2 | 625 | 23 | AAO21862 |
| 16 | 1776   | 55.2 | 637 | 23 | AAO21861 |
| 17 | 1774   | 55.2 | 637 | 24 | AAU77532 |
| 18 | 1620.5 | 50.4 | 628 | 23 | AAO21847 |
| 19 | 1532.5 | 47.7 | 641 | 23 | AAO21845 |
| 20 | 1504.5 | 46.8 | 648 | 20 | AA52834  |
| 21 | 1491.5 | 46.4 | 648 | 23 | AAO21856 |
| 22 | 1485.5 | 46.2 | 720 | 21 | AAV97422 |
| 23 | 1442   | 44.8 | 640 | 23 | AAO21858 |
| 24 | 1439   | 44.7 | 636 | 23 | AAO21857 |
| 25 | 1420.5 | 44.2 | 708 | 21 | AAV97415 |
| 26 | 1418.5 | 44.1 | 724 | 20 | AAV14145 |
| 27 | 1418.5 | 44.1 | 727 | 20 | AAV14146 |
| 28 | 1411.5 | 43.9 | 637 | 23 | AAO21855 |
| 29 | 1411.5 | 43.9 | 659 | 23 | AAU48246 |
| 30 | 1411.5 | 43.9 | 717 | 21 | AAU42534 |
| 31 | 1411.5 | 43.9 | 717 | 23 | AAU48245 |
| 32 | 1411.5 | 43.9 | 824 | 23 | AAU48247 |
| 33 | 1410.5 | 43.9 | 641 | 23 | AAO21854 |
| 34 | 1410.5 | 43.9 | 721 | 21 | AAV97414 |
| 35 | 1407.5 | 43.8 | 725 | 21 | AAV15157 |
| 36 | 1406.5 | 43.7 | 637 | 20 | AAV52833 |
| 37 | 1404.5 | 43.7 | 717 | 21 | AAV51612 |
| 38 | 1404.5 | 43.7 | 717 | 21 | AAV51667 |
| 39 | 1402.5 | 43.6 | 719 | 23 | AAO21851 |
| 40 | 1402   | 43.6 | 735 | 23 | AAO21849 |
| 41 | 1394.5 | 43.4 | 633 | 21 | AAU20542 |
| 42 | 1394.5 | 43.4 | 633 | 21 | AAU20543 |
| 43 | 1394.5 | 43.4 | 633 | 23 | AAO21848 |
| 44 | 1392   | 43.3 | 721 | 20 | AAV14144 |
| 45 | 1389.5 | 43.2 | 632 | 21 | AAU42535 |

#### ALIGNMENTS

RESULT 1  
ID ABG61581 standard; Protein: 620 AA.  
XX  
AC ABG61581;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DB High growth methanotrophic bacterial strain polypeptide #31.  
XX  
XX High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme;  
XX methane; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;  
XX pyrophosphate dependent phosphofructokinase; nitrogen-containing compound;  
XX ammonias; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;  
XX methane-containing environment; waste water treatment system; isoprenoid;  
XX nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.  
XX  
XX Methylobionas 16a.  
XX  
XX MO200220728-A2.  
XX  
XX  
XX PD 14-MAR-2002.  
XX  
XX 28-AUG-2001; 2001WO-US26827.  
XX PF  
XX PR 01-SEP-2000; 2000US-229858P.  
XX  
XX (DUPO) DU POINT DE MEMOIRS & CO E. I.  
XX  
XX Koffas M, Odom JM, Schenkele A;  
XX  
XX WPI; 2002-452200/48.

DR N-PSDB; ABEK3260.

XX New high growth methanotrophic bacterial strain, useful for producing  
PT single cell proteins, grows on a C1 carbon substrate, and comprises a  
PT functional gene encoding in Embden-Meyerhof carbon pathway  
PS Claim 11; Page 138-140; 157pp; English.

XX The invention relates to a high growth methanotrophic bacterial strain,  
XX which grows on a C1 carbon substrate e.g. methane and methanol, and  
CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a  
CC gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S  
CC rRNA. The bacterial strain is useful for the production of single cell  
CC protein and for the biotransformation of a nitrogen-containing compound,  
CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the  
CC production of a feed product comprising a protein, carbohydrates and a  
CC pigment and for reducing oxygen demand, for removing nitrates and  
CC nitrates in methane-containing environments such as landfills, waste  
CC water treatment systems or anywhere that methane, oxygen and nitrates are  
CC present. The bacterial strain of the invention can be used as a  
CC denitrifying agent for the conversion of nitrate or nitrite to nitrous  
CC oxide with methane or methanol as a carbon source. It is also used in the  
CC production of biomass including proteins, carbohydrates and a wide  
CC variety of pigments (particularly for isoprenoid pigments for the  
CC purpose of generating animal feeds), in production of terpenoid and  
CC carotenoid compounds, useful as pigments and as monomers in polymeric  
CC materials and in production of exopolysaccharides at high levels.  
CC Sequences ABE61551-ABE61590 represent high growth methanotrophic  
CC bacterial strain proteins of the invention.

XX Sequence 620 AA;

XX Query Match 100.0%; Score 3216; DB 23; Length 620;

XX Best Local Similarity 100.0%; Pred. No. 1.2e-285; Indels 0; Gaps 0;

XX Matches 620; Conservative 0; Mismatches 0;

DB 1 MKLTTDYPPLKNTHTPADIRALSKDQLQDLADVRGYLTHTVTSISGHPFAGIGTVELTV 60

QY 61 ALHYVPTPVQDLVWDVGHQAYPHKILITGRKEMPTIRTCGYSAFAPABESYDAFGVG 120

DB 61 ALHYVPTPVQDLVWDVGHQAYPHKILITGRKEMPTIRTCGYSAFAPABESYDAFGVG 120

QY 121 HSTSTISALGMAIASOLRGEDKKMAVIGDGSITGGAYEAMNADGVANILVILNDN 180

DB 121 HSTSTISALGMAIASOLRGEDKKMAVIGDGSITGGAYEAMNADGVANILVILNDN 180

QY 181 DMSISPPVGMANNVLTLYLSKSYSSVRESKKAIAAPSVWELARKTEHHVKATYRGCT 240

DB 181 DMSISPPVGMANNVLTLYLSKSYSSVRESKKAIAAPSVWELARKTEHHVKATYRGCT 240

QY 241 LFEELGRVYRPPIDGHVEMVSTLENTKDGTPVPLAVTKKGGAPAKEDPLAHGV 300

DB 241 LFEELGRVYRPPIDGHVEMVSTLENTKDGTPVPLAVTKKGGAPAKEDPLAHGV 300

QY 301 PAPPTKDFLKAAPSHPITYTEVGRWLCDMAADERLIGITPAMEGSGIVFESQKFP 360

DB 301 PAPPTKDFLKAAPSHPITYTEVGRWLCDMAADERLIGITPAMEGSGIVFESQKFP 360

QY 361 NRYFDVAIAEGHANTVLAAGCCACAKPVVATYSTFLQNGNQLHDVALQNLMDLFLADR 420

DB 361 NRYFDVAIAEGHANTVLAAGCCACAKPVVATYSTFLQNGNQLHDVALQNLMDLFLADR 420

QY 421 AGLVGPDPTAGAFDYSYRCIPNMLIMADENECROMLTGFOHHGPASVYPRGKG 480

DB 421 AGLVGPDPTAGAFDYSYRCIPNMLIMADENECROMLTGFOHHGPASVYPRGKG 480

QY 481 PGAAIDPFLTALIGKAEVRHHSRIALILANGSVTPPVEGKQLGAVVNMRRVKKPDQ 540

DB 481 PGAAIDPFLTALIGKAEVRHHSRIALILANGSVTPPVEGKQLGAVVNMRRVKKPDQ 540

QY 541 ALVIELARTHDVFTVEENVIAGAGSALNTFLQAKVILMPVCNIGLDFRVEQGSREEL 600

DB 541 ALVIELARTHDVFTVEENVIAGAGSALNTFLQAKVILMPVCNIGLDFRVEQGSREEL 600

QY 601 LSLVGLDSKGLATTEQFCA 620  
DB 601 LSLVGLDSKGLATTEQFCA 620

RESULT 2

AAE22301  
ID AAE22301 standard; Protein; 620 AA.

AC AAE22301;

DE 25-TUL-2002 (first entry)

XX Methylomonas 16a sp. D-1-deoxyxylulose-5-phosphate synthase (Dxs) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; aetaxanthin; diet;

XX anti-oxidant; steroid; flavour; fragrance; electro-optic applications;

XX aquaculture; enzyme; D-1-deoxyxylulose-5-phosphate synthase; Dxs.

XX Methylomonas 16a sp.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US27420.

XX 01-SEP-2000; 2000US-229858P.

XX 01-SEP-2000; 2000US-229907P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Brostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odom JM, Ficataggio SK, Rouviere PE;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35499.

XX Claim 41; Page 109-111; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds.

XX The method comprises a transformed metabolising host cell, comprising

XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

XX encoding an enzyme in the carotenoid biosynthetic pathway, under the

XX control of regulatory sequences, and contacting the host cell with carbon

XX substrate to produce a carotenoid compound. The method is useful for

XX producing carotenoid compounds such as antheraxanthin and aetaxanthin, by

XX using microorganism having a nucleic acid molecule encoding enzymes in

XX the carotenoid biosynthetic pathway and which metabolize single carbon

XX substrates. The carotenoids have potent anti-oxidant properties useful in

XX diet, and aquaculture elements. The carotenoids are also useful as

XX intermediates in the synthesis of steroids flavours and fragrances and

XX compounds for potential electro-optic applications. The present sequence

XX is Methylomonas 16a sp. D-1-deoxyxylulose-5-phosphate synthase (Dxs)

XX enzyme used in the invention.

XX Sequence 620 AA;

XX Query Match 100.0%; Score 3216; DB 23; Length 620;

XX Best Local Similarity 100.0%; Pred. No. 1.2e-285; Indels 0; Gaps 0;

XX Matches 620; Conservative 0; Mismatches 0;

DB 1 MKLTTDYPPLKNTHTPADIRALSKDQLQDLADVRGYLTHTVTSISGHPFAGIGTVELTV 60

QY 1 MKLTTDYPPLKNTHTPADIRALSKDQLQDLADVRGYLTHTVTSISGHPFAGIGTVELTV 60

QY 61 ALHYVNTPVQDQVMDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120  
 DB 61 ALHYVNTPVQDQVMDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120  
 QY 121 HSTSTISALGMAIASQLRGEDKRWVAIIGDGSITGGMAVEANAGVNNALVITLNDN 180  
 DB 121 HSTSTISALGMAIASQLRGEDKRWVAIIGDGSITGGMAVEANAGVNNALVITLNDN 180  
 QY 181 DMSISPPVGMANNITLTKVLSKFTSSVREBSKKALAMPVWELARKTEEHVKMTVPQT 240  
 DB 181 DMSISPPVGMANNITLTKVLSKFTSSVREBSKKALAMPVWELARKTEEHVKMTVPQT 240  
 QY 241 LPEELGFNFEGPIIDGDMVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300  
 DB 241 LPEELGFNFEGPIIDGDMVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300  
 QY 301 PAFDPTKDFLEPKAAPSPHPTYTEVFGRLCDMAADDERLLGITPAMREGSLVPSQKFP 360  
 DB 301 PAFDPTKDFLEPKAAPSPHPTYTEVFGRLCDMAADDERLLGITPAMREGSLVPSQKFP 360  
 QY 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRGYDQIHDAVALQNDMLFALDR 420  
 DB 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRGYDQIHDAVALQNDMLFALDR 420  
 QY 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFQHHGPASVRYPRGK 480  
 DB 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFQHHGPASVRYPRGK 480  
 QY 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMVTTPAVEAGKQIGATVNNRFXKPPDQ 540  
 DB 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMVTTPAVEAGKQIGATVNNRFXKPPDQ 540  
 QY 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIGLPDRFVEQGSREEL 600  
 DB 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIGLPDRFVEQGSREEL 600  
 QY 601 LSLVGLDSKGLIATIEQFCA 620  
 DB 601 LSLVGLDSKGLIATIEQFCA 620

RESULT 3

AAU80325  
 ID AAU80325 standard; Protein; 620 AA.  
 XX AC  
 XX AAU80325;  
 XX DT 15-JUN-2002 (first entry)  
 XX DE Methylomonas 16a ORF1 dks (1-deoxyxylulose-5-phosphate synthase) protein.  
 XX KM Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
 XX KM kerateneid; pigment; flavour; fragrance; open reading frame 1; ORF1.  
 XX KM dks; 1-deoxyxylulose-5-phosphate synthase enzyme.  
 XX OS Methylomonas sp.  
 XX PN WO200220733-A2.  
 XX PD 14-MAR-2002.  
 XX PF 29-AUG-2001; 2001WO-US26852.  
 XX PR 01-SEP-2000; 2000US-229907P.  
 XX PA (DUPO) DU PONT DE NEMOURS & CO E I.  
 PI Cheng Q, Koffas M, Norton KC, Odom JM, Picataggio SK, Rouviere PB;  
 PI Schenzzle A, Tomb J;  
 DR WPI; 2002-383051/41.  
 DR N-PSDB; AERK50081.

XX Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,  
 PT isolated from Methylomonas 16a, useful for the production of isoprenoid  
 PT compounds -  
 XX  
 XX PS  
 XX Claim 4; Page 66-68; 84pp; English.  
 CC The present invention relates to a new nucleic acid molecule encoding  
 CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.  
 CC The invention is useful for obtaining a nucleic acid molecule  
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the  
 CC microbial production of isoprenoid compounds. The molecules of the  
 CC invention are also useful for regulating isoprenoid biosynthesis in an  
 CC organism and for producing recombinant organisms for producing various  
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,  
 CC for the production of kerateneids and their derivatives, isoprenoid  
 CC intermediates, and as pure products useful as pigments, flavours and  
 CC fragrances. The present amino acid sequence represents the Methylomonas  
 CC 16a open reading frame 1 (ORF1) dks (1-deoxyxylulose-5-phosphate synthase  
 CC enzyme) protein of the invention, as described above.  
 XX  
 XX Sequence 620 AA;  
 Query Match 100.0%; Score 3216; DB 23; Length 620;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-285;  
 Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLITDYPILANKITPADIRALSKDQLOQLADVRGTLTHVVISGGHFAAGLGLVELTV 60  
 DB 1 MKLITDYPILANKITPADIRALSKDQLOQLADVRGTLTHVVISGGHFAAGLGLVELTV 60  
 QY 61 ALHYVNTPVQDQVMDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120  
 DB 61 ALHYVNTPVQDQVMDVGHQAYPHKILTGKREMPRTITLGGVSAFPARDSEYDAFGV 120  
 QY 121 HSTSTISALGMAIASQLRGEDKRWVAIIGDGSITGGMAVEANAGVNNALVITLNDN 180  
 DB 121 HSTSTISALGMAIASQLRGEDKRWVAIIGDGSITGGMAVEANAGVNNALVITLNDN 180  
 QY 181 DMSISPPVGMANNITLTKVLSKFTSSVREBSKKALAMPVWELARKTEEHVKMTVPQT 240  
 DB 181 DMSISPPVGMANNITLTKVLSKFTSSVREBSKKALAMPVWELARKTEEHVKMTVPQT 240  
 QY 241 LPEELGFNFEGPIIDGDMVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300  
 DB 241 LPEELGFNFEGPIIDGDMVEMLVSTLENLKDLPVFLHVYTKKGYAPAEKOPLAYHGV 300  
 QY 301 PAFDPTKDFLEPKAAPSPHPTYTEVFGRLCDMAADDERLLGITPAMREGSLVPSQKFP 360  
 DB 301 PAFDPTKDFLEPKAAPSPHPTYTEVFGRLCDMAADDERLLGITPAMREGSLVPSQKFP 360  
 QY 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRGYDQIHDAVALQNDMLFALDR 420  
 DB 361 NRYFDVAIAEQHAYVTLAAGACQAKPVVAISTFLQRGYDQIHDAVALQNDMLFALDR 420  
 QY 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFQHHGPASVRYPRGK 480  
 DB 421 AGLVGPDGFTTHAGAFDYSYMRCTPNNLIMAPADENECRQMLTTGFQHHGPASVRYPRGK 480  
 QY 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMVTTPAVEAGKQIGATVNNRFXKPPDQ 540  
 DB 481 PGAAIDPITLALBEGKAEVREHGSRIAILAMGSMVTTPAVEAGKQIGATVNNRFXKPPDQ 540  
 QY 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIGLPDRFVEQGSREEL 600  
 DB 541 ALVLELARTHDVFTVEENVVLAGAGSAINTEFLQAKVLMPCVNIGLPDRFVEQGSREEL 600  
 QY 601 LSLVGLDSKGLIATIEQFCA 620  
 DB 601 LSLVGLDSKGLIATIEQFCA 620

RESULT 4

AA00012  
ID AA00012 standard; Protein, 627 AA.  
XX  
XX AA00012;  
AC  
XX 08-MAY-2001 (first entry)  
XX  
XX Pseudomonas aeruginosa Deoxyxylulose-5-phosphate synthase, dxs.  
DE  
XX Deoxyxylulose-5-phosphate synthase; dxs; agonist; antagonist;  
XX diagnostic; eukaryotic extracellular matrix protein; antibody; wound;  
XX in-dwelling device; Helicobacter pylori infection; gastric ulcer;  
XX gastritis; gastrointestinal carcinoma.  
XX  
XX Pseudomonas aeruginosa.  
OS  
XX  
XX WO200114409-A1.  
XX  
XX 01-MAR-2001.  
XX  
XX 16-AUG-2000; 2000WO-US22332.  
XX  
XX 19-AUG-1999; 99US-0377279.  
XX  
XX (SMK ) SMITHKLINE BEECHAM CORP.  
XX (SMK ) SMITHKLINE BEECHAM PLC.  
XX  
XX Huang J, Jiang X, McDevitt D;  
XX  
XX WPI; 2001-211303/21.  
XX  
XX N-PSDB; AAS00017.  
XX  
XX New deoxyxylulose-5-phosphate (dxs) polypeptides of Pseudomonas  
PT aeruginosa useful for diagnosis of dxs expression or activity related  
PT disease -  
XX  
XX Claim 1; Page 4; 41pp; English.  
XX  
XX The sequence represents Pseudomonas aeruginosa deoxyxylulose-5-phosphate  
CC synthase, dxs. (Ant)agonists of dxs are useful for treating an individual  
CC in need of enhanced or inhibited activity or expression of or  
CC immunological response to dxs. Dxs and the nucleic acid encoding it are  
CC useful as diagnostic reagents, especially in relation to diseases related  
CC to expression or activity of dxs. Fragments of the nucleic acid are  
CC useful as probes or primers and to synthesize full length dxs  
CC polynucleotides. The nucleic acid and protein are useful for diagnosing  
CC bacterial infections, especially by Pseudomonas aeruginosa, staging a  
CC disease or response of an infectious organism to drugs, and are also used  
CC in a structure based design of an (ant)agonist to dxs. The nucleic acid,  
CC protein and antibodies raised against it are useful for screening for  
CC (ant)agonists of dxs or the nucleic acid and are useful to interfere with  
CC the initial physical interaction between a pathogen and a mammalian host  
CC responsible for sequelae of infection. These molecules are useful in  
CC preventing adhesion of gram positive and/or gram negative bacteria to  
CC eukaryotic extracellular matrix proteins or in-dwelling devices or in  
CC wounds; to block bacterial adhesion; to block the normal progression of  
CC pathogenesis in infections initiated other than by the implantation of  
CC in-dwelling devices or by other surgical techniques. Dxs is also useful  
CC for identifying membrane bound or soluble receptors. The (ant)agonists  
CC are useful for treating Helicobacter pylori infection, and also to  
CC prevent, inhibit or treat gastric ulcers, gastritis and gastrointestinal  
CC carcinoma.  
XX  
XX  
XX Sequence 627 AA;  
SQ

Query Match 64.4%; Score 209.5; DB 22; Length 627;  
Best Local Similarity 63.4%; Pred. No. 1.7e-180;  
Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;

QY 8 PLKRIHTPADRALSKDOLADLVRCGLTHYVSGSGHPAALGTVELTVALHYEN 67  
DB 16 PLDRASSPAERLRGAEDELFTLADLRLQVLLYTVGQGGHFGAGLGVVELTIALHYED 75

QY 68 TPVDQLVWDVGHQAVPEKILTGRKERMPITITLGGVSAFPARDSESYDAFGVGHSTIS 127  
DB 76 TPDDRLLVWDVGHQAVPEKILTGRKERMPITITLGGVSAFPARDSESYDAFGVGHSTIS 135  
QY 128 AALGKALASQLRGDKQWVAVIIGDSITGCAVEAMONHAYVNNLVLTLNDNMSTSP 187  
DB 136 AALGKALASQLRGDKQWVAVIIGDSITGCAVEAMONHAYVNNLVLTLNDNMSTSP 195  
QY 188 VQAMNNVITVYLSKFFSVSVEESKKALAKMPVWEELARKTEEHVKMIVPGLPEELGF 247  
DB 196 VGSLSNVLAKLISRTYSMSKESKVLRLPGAMEIARTEEYAKGLVPGTLPEELGW 255  
QY 248 NYRPIGDCHDVEMVSTLENLKDITGPVPLVWTKKKGVA PAKKDLAVHGVAFDPTX 307  
DB 256 NYRPIGDCHDLPTVATLNNRDKKQPFHVVTKKGFAPALDPTGYHAI ---TK 310  
QY 308 DFLPAAFP--SPHPTYTEVFGRMLCDMAADERLLGITPARREGSGVFEQKPNRYFD 365  
DB 311 LEAPGSAFKKTGGKPSVFGOMLCDMAQDARILGITPMKESGDLVAFSERPERYFD 370  
QY 366 VAIABQHAVTTLAQAQACQAKPVAIYSTIQRGIDQLIHDAIOMLDMALPABRAGIV 425  
DB 371 VAIABQHAVTTLAQAQACQAKPVAIYSTIQRGIDQLIHDAIOMLDMALPABRAGIV 430  
QY 426 PDGPTHGAPFYSYMRCPNMLIMAPADENCRQMLTTGFOHGPASVYPRGKPGAAI 485  
DB 431 EDGPTHGAPFYSYMRCPNMLIMAPADENCRQMLTTGFOHGPASVYPRGKPGAAI 490  
QY 486 DFTLTALRIGKAEVRRHGSRIATLAWGSMVTPAVEACKQLGATVANNRFPVFPQALVLE 545  
DB 491 DPDLQPEIGKGVRRGRGVALLVFGVQLAEAKRVASLDATVAVDMRFVFPDLDEALVRE 550  
QY 546 LARTRHDVFWVEENVVIGAGSAINTELQAKVLMPCNIGLPDRFVQSGREELISIVG 605  
DB 551 LAGSHLVTTEENAVWGAGSAGVETLASBGLVPLQLGELPDYVYHAKPSTMLAECG 610  
QY 606 LDSKGIATTEQ 617  
DB 611 IDAAGIRKAVRQ 622

RESULT 5  
AA021846  
ID AA021846 standard; Protein, 627 AA.  
XX  
XX AA021846;  
AC  
XX 13-SEP-2002 (first entry)  
DT  
XX  
XX Isoprenoid related protein sequence SEQ ID No 15.  
DE  
XX  
XX Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
XX decaprenyl diphosphate synthase.  
KM  
XX Pseudomonas aeruginosa.  
OS  
XX  
XX WO200226933-A2.  
XX  
XX 04-APR-2002.  
XX  
XX 28-SEP-2001; 2001WO-US30328.  
XX  
XX 29-SEP-2000; 2000US-236580P.  
XX  
XX (CRGI ) CARGILL INC.  
XX  
XX Gokarn R, Jessen H, Zidwick MJ;  
XX WPI; 2002-416486/44.  
XX  
XX Substantially pure polypeptides having e.g.,  
PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
PT production of isoprenoids, especially CoQ(10) -



XX  
PS Disclosure; Fig 6; 246pp; English  
vv

The invention relates to methods and materials for the production of isoprenoids. More particularly the invention provides isolated nucleic acids, substantially pure polypeptides, host cells, and methods for producing various isoprenoid compounds. The polypeptides are useful for the production of isoprenoids, especially Coq(10). Expressing the pure polypeptides, which has 1-oxoheptylulose-5-phosphate synthase (DXS) activity or decaprenyl diphosphate synthase (DPS) activity, is useful for increasing production of Coq(10) in a cell having endogenous DXS activity. This sequence represents a protein relating to the isoprenoid production of the invention.

Sequence 627 AA; SQ

|                       |                  |                    |          |            |
|-----------------------|------------------|--------------------|----------|------------|
| Query Match           | 64.4%            | Score 2059.5       | DB 23    | Length 627 |
| Best Local Similarity | 63.4%            | Pred. No. 1.7e-180 |          |            |
| Matches 388           | Conservative 100 | Mismatches 117     | Indels 7 | Gaps 2     |

|    |     |                  |                    |                |                |     |
|----|-----|------------------|--------------------|----------------|----------------|-----|
| QY | 8   | PLIXNHTEPADIDALS | KDOLLOOLABVRYLTHYS | ISGCHPAAIGVVEL | VALVHYEN       | 67  |
| Db | 16  | PLIDRRASSPHEL    | RRLGRADLET         | LABELRYLLYLT   | YQIGCHGAGVVELT | 75  |
| QY | 68  | TPVDOLVMDVGHQ    | AYHKKLTGCRKEMPT    | TIRTLGVSAP     | PAADSEYDA      | 12  |
| Db | 76  | TPDRLVMDVGHQ     | AYHKKLTERRRIMGL    | TIRKXNGLA      | PPRAAE         | 135 |
| QY | 128 | AALGAMLA         | BQLREDEKDKVA       | IAIGDSIT       | TGMAVEAMNAG    | 18  |
| Db | 136 | AALGAMLA         | ARLOCEKREKS        | VA             | AVIGDAL        | 196 |
| QY | 188 | VGMANNVYL        | KSXKSVSREESK       | KALAKMPSV      | VELAKRTB       | 24  |
| Db | 196 | VGGSLN           | LAKLISSTRYS        | SMRBSKKV       | SRLLGAW        | 255 |
| QY | 248 | NYGP             | PLDGEVMDL          | YSTLENL        | DLTGVFL        | 307 |
| Db | 256 | NYGP             | PLDGEHDL           | PLTVAT         | LNMEDM         | 310 |
| QY | 308 | DLEPKAP          | --SHP              | PTYEEV         | FGFRLCDMA      | 365 |
| Db | 311 | LEAKS            | APKKTGPK           | KYSV           | FGQMLCDMA      | 370 |
| QY | 366 | VAIABGH          | AVTLAAGOC          | QCAKPV         | VALYST         | 425 |
| Db | 371 | VAIAEGH          | AVTLAAGAC          | CEMKRP         | VAIYST         | 430 |
| QY | 426 | PGGPT            | AGAFDYS            | YNKRCIP        | PNMLIMAP       | 485 |
| Db | 431 | EGGPT            | AGASFD             | SYLARCIP       | GMVMP          | 490 |
| QY | 486 | DEPL             | TLALETG            | KEVHHG         | SRILAILAM      | 545 |
| Db | 491 | DEPL             | CPVEIG             | KVNVRR         | GGRVALLV       | 550 |
| QY | 546 | LARTD            | VPFVTE             | VENYIAG        | AGSAIT         | 605 |
| Db | 551 | LAGS             | HELAVLT            | IBENNV         | MGASGA         | 610 |
| QY | 606 | LD               | SKGLAT             | IEQ            |                |     |
| Db | 611 | LD               | AAIG               | IERA           | VR             |     |

DE *Pseudomonas aeruginosa* dxs ts-80 mutant.

XX Deoxyxylulose-5-phosphate synthase; dxs; agonist; antagonist;  
XX diagnostic; eukaryotic extracellular matrix protein; antibody; wound  
XX in-dwelling device; *Helicobacter pylori* infection; gastric ulcer;  
XX gastritis; gastrointestinal carcinoma; mutant; ts-32.  
OS *Pseudomonas aeruginosa*.

*Pseudomonas aeruginosa*.

|    |                 |  |
|----|-----------------|--|
| PH | Key             | Location/Qualifiers                    |
| FT | Misc-difference | 380                                    |
| FT |                 | /note= "Wild-type Thr replaced by Ile" |
| xx |                 |  |

PN WO200114409-A1

PD 01-MAR-2001.

PF 16-AUG-2000; 2000WD-US22332.

PR 19-AUG-1999; 99US-0377279.

PA (SMIK ) SMITHKLINE BEECHAM CORP.

(SMIK ) SMITHKLINE BEECHAM CORP.

PI Huang J, Jiang X, McDevitt D;

DR WPI; 2001-211303/21.

DR N-PSDB; AAAS00019.

PT New deoxyxylulose-5-phosphate (dxs) polypeptides of *Pseudomonas*  
PT aeruginosa useful for diagnosis of dxs expression or activity related  
PT disease -

PS Example 2; Page - ; 41pp; English.

The sequence represents *Pseudomonas aeruginosa* deoxyribonucleic acid synthase, dxs, temperature sensitive mutant, ts-80. (Ant)agonists of dxs are useful for treating an individual in need of enhanced or inhibited activity or expression of or immunological response to dxs. Dxs and the nucleic acid encoding it are useful as diagnostic reagents, especially in relation to diseases related to expression or activity of dxs. Fragments of the nucleic acid are useful as probes or primers and to synthesise full length dxs polynucleotides. The nucleic acid and protein are useful for diagnosing bacterial infections, especially by *Pseudomonas aeruginosa*, straining a disease or response of an infectious organism to drugs, and are also used in a structure based design of an (ant)agonist to dxs. The nucleic acid, protein and antibodies raised against it are useful for screening for (ant)agonists of dxs or the nucleic acid and are useful to interfere with the initial physical interaction between a pathogen and a mammalian host responsible for sequelae of infection. These molecules are useful in preventing adhesion of gram positive and/or gram negative bacteria to eukaryotic extracellular matrix proteins or in-dwelling devices or in wounds; to block bacterial adhesion; to block the normal progression of pathogens in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques. Dxs is also useful for identifying membrane bound or soluble receptors. The (ant)agonists are useful for treating *Helicobacter pylori* infection, and also to prevent, inhibit or treat gastric ulcers, gastritis and gastrointestinal carcinoma.

The present sequence is not shown in the specification but is derived from the dxs sequence given in AAU00012.

The present sequence is not shown in the specification but is derived from the dms sequence given in A400012.

Sequence 627 AA;

|                       |                   |                     |           |             |
|-----------------------|-------------------|---------------------|-----------|-------------|
| Query Match           | 64.2%;            | Score 2063.5;       | DB 22;    | Length 627; |
| Best Local Similarity | 63.2%;            | Pred. No. 6.1e-180; |           |             |
| Matches 387;          | Conservative 100; | Mismatches 118;     | Indels 7; | Gaps 2      |

**QY**   8 P L I K N I H F P A D I R A S T Q U L O L A D E V R G Y / E H V S I S G G F P A G C G T V E L T V A L A V E N     67  
     | | | : : : : : | : : : : : | : : : : :  
**Db**   16 P L I D B A S S P A E R L G E A D E T L A D E R Q U M L Y A V G O T G G H F G A G W E L T M A T V E D     75  
            | : : : : : | : : : : : | : : : : :  
**OY**   68 T P V D Q L V M D V G H Q A V P H K I L I N G R E K R M P T I R T L G G V S A F P A R D E S B Y D A R P V G H S T S I S   127  
                | : : : : : | : : : : : | : : : : :

```

Db 76 TPDDRVLVMDVGHQAPPHKILTERELMGLRQKNGIAAPRRASEHYDFRGHSSTIS 135
Qy 128 AALGMAIASQLGEGDKKQVAIIGDGSITGMAVEANNAADVANNILVINDMSISPP 187
Db 136 AALGMAIAARLOQEKERSVAIVGDGALTGMAFEALNHASEVDADMEVILINDMSISHN 195
Qy 188 VGAMNNYLTKVLSKREYSSVREESKKALAMPVWELARKTEEHYKGMIVPGTLFEEIGF 247
Db 196 VGLSNVTLAKILSSRYSSMRGSKKYLRLPGAMBIARTTEYAKKGLVPGTLFEEIGM 255
Qy 248 NYFGPIDGHDVEMLVSTLENLKDLPVFLVYVYKKGAVAPAEKDPPLAYHGVPAFDPK 307
Db 256 NYIGPIDGHDLPVLVATLNNRDMKGPFLHVVTKKGKFAPELDPFGYHAI-----TK 310
Qy 308 DFLPKAAP--SPHPTTYVGRMLCDMAADERLIGITPAMREGSLVFEQKXPNRYPD 365
Db 311 LEAPGAPKKTGGPKTSVFGQMLCDMAADARLLGITPAMREGSDLVAFSERYERYPD 370
Qy 366 VALAEQHAVTLAAGQACQAKPEVVAIYSTFLQRYDOLIHVYALQNLDMPLALDRAGLVG 425
Db 371 VALAEQHAVTLAAGMACGEMKPEVVAIYSTFLQRAYDOLIHVAVQHLDLVFLAIDRAGLVG 430
Qy 426 PGGPTTAGAFDYSYKACIPMNLIMAPADNECQMLTTFQFHGPASTRYPRGKPGAAI 485
Db 431 ECGPTTAGSFDISYKCIPLGMLWMPSPDEDLRKLITGVLPDGAAYRYPGSGPNHPI 490
Qy 486 DFLTALAEIGKAEVHHGSRILIAMGWSVTPAVEAGKQLGATVNNKAFVFPDQALVLE 545
Db 491 DDLDQVEIGKGVNRGRGRVALLVFGVGLAEMAMVASLDATVYDFRFPDLBALVRE 550
Qy 546 LARTDVPVTVTEENIAGAGSALNTFLQAGVTLMPVNCIGLPDRFVEQGSREBLISVG 605
Db 551 IAGSHELLVLTIEENVMVGAGSAGVEFLASEGLVEPLLQGLPDYVYVHAKPSEMLAEQG 610
Qy 606 LPSKGLATIEQ 617
Db 611 LDAGIGERAVRQ 622

```

RESULT 7  
AAU00013  
ID AAU00013 standard; Protein; 627 AA.

XX AAU00013;

DT 08-MAY-2001 (first entry)

DE Pseudomonas aeruginosa dxs ts-32 mutant.

KM Deoxyxylulose-5-phosphate synthase; dxs; agonist; antagonist;  
diagnostic; eukaryotic extracellular matrix protein; antibody; wound;  
in-dwelling device; Helicobacter pylori infection; gastric ulcer;  
gastritis; gastrointestinal carcinoma; mutant; mutagen; ts-32.

OS Pseudomonas aeruginosa.

XX Key Location/Qualifiers

FT Misc-difference 115 /note= "Wild-type Pro replaced by Ser"

FN MO200114409-A1.

PD 01-MAR-2001.

PF 16-AUG-2000; 2000MO-US22332.

PR 19-AUG-1999; 99US-0377279.

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PI Huang J, Jiang X, McDevitt D;

XX WPI. 2001-211303/21.  
DR N-PDB; AAS00018.

PT New deoxyxylulose-5-phosphate (dxs) polypeptides of Pseudomonas  
aeruginosa useful for diagnosis of dxs expression or activity related  
disease

PS Example 2; Page -; 41pp; English.

The sequence represents Pseudomonas aeruginosa deoxyxylulose-5-phosphate  
synthase, dxs, temperature sensitive mutant, ts-32. (Ant)agonists of dxs  
are useful for treating an individual in need of enhanced or inhibited  
activity or expression of or immunological response to dxs. Dxs and the  
nucleic acid encoding it are useful as diagnostic reagents, especially in  
relation to diseases related to expression or activity of dxs. Fragments  
of the nucleic acid are useful as probes or primers and to synthesize  
full length dxs polynucleotides. The nucleic acid and protein are useful  
for diagnosing bacterial infections, especially by Pseudomonas  
aeruginosa, staging a disease or response of an infectious organism to  
drugs, and are also used in a structure based design of an (ant)agonist  
to dxs. The nucleic acid, protein and antibodies raised against it are  
useful for screening for (ant)agonists of dxs or the nucleic acid and are  
useful to interfere with the initial physical interaction between a  
pathogen and a mammalian host responsible for sequelae of infection.  
These molecules are useful in preventing adhesion of gram positive and/or  
gram negative bacteria to eukaryotic extracellular matrix proteins or  
in-dwelling devices or in wounds, to block bacterial adhesion; to block  
the normal progression of pathogenesis in infections initiated other than  
by the implantation of in-dwelling devices or by other surgical  
techniques. Dxs is also useful for identifying membrane bound or soluble  
receptors. The (ant)agonists are useful for treating Helicobacter pylori  
infection, and also to prevent, inhibit or treat gastric ulcers,  
gastritis and gastrointestinal carcinoma.  
The present sequence is not shown in the specification but is derived  
from the dxs sequence given in AAU00012.

CC Sequence 627 AA;

Query Match 64.1%; Score 2061.5; DB 22; Length 627;  
Best Local Similarity 63.2%; Pred. No. 9.3e-180;  
Matches 387; Conservative 100; Mismatches 118; Indels 7; Gaps 2;

```

Qy 8 PLKNIHTPADIALSKDQLOGLADRYGTLTHTVISGCHFAAGLCTVLTALHYVN 67
Db 16 PLDRASSPAELRLTGADLELRLQYLTIVQGTGHPAGLGVLLTALHAYVD 75
Qy 68 TPVDQVLVMDVGHQAPPHKILTERELMGLRQKNGIAAPRRASEHYDFRGHSSTIS 127
Db 76 TPDDRVLVMDVGHQAPPHKILTERELMGLRQKNGIAAPRRASEHYDFRGHSSTIS 135
Qy 128 AALGMAIASQLGEGDKKQVAIIGDGSITGMAVEANNAADVANNILVINDMSISPP 187
Db 136 AALGMAIAARLOQEKERSVAIVGDGALTGMAFEALNHASEVDADMEVILINDMSISHN 195
Qy 188 VGAMNNYLTKVLSKREYSSVREESKKALAMPVWELARKTEEHYKGMIVPGTLFEEIGF 247
Db 196 VGLSNVTLAKILSSRYSSMRGSKKYLRLPGAMBIARTTEYAKKGLVPGTLFEEIGM 255
Qy 248 NYFGPIDGHDVEMLVSTLENLKDLPVFLVYVYKKGAVAPAEKDPPLAYHGVPAFDPK 307
Db 256 NYIGPIDGHDLPVLVATLNNRDMKGPFLHVVTKKGKFAPELDPFGYHAI-----TK 310
Qy 308 DFLPKAAP--SPHPTTYVGRMLCDMAADERLIGITPAMREGSLVFEQKXPNRYPD 365
Db 311 LEAPGAPKKTGGPKTSVFGQMLCDMAADARLLGITPAMREGSDLVAFSERYERYPD 370
Qy 366 VALAEQHAVTLAAGQACQAKPEVVAIYSTFLQRYDOLIHVYALQNLDMPLALDRAGLVG 425
Db 371 VALAEQHAVTLAAGMACGEMKPEVVAIYSTFLQRAYDOLIHVAVQHLDLVFLAIDRAGLVG 430
Qy 426 PGGPTTAGAFDYSYKACIPMNLIMAPADNECQMLTTFQFHGPASTRYPRGKPGAAI 485

```

Db 431 EDGPTAGSFDISYACIPQMLVMPSPDDEDEKLLTGYLFDGPAVAVPRGSGPNHPI 490  
 QY 486 DPLTLALTEGKAEVRHSGRIALANGSNVTPAVEKQKQATVVMRRVPPDQALVLE 545  
 Db 491 DPLQGEVEIGKGVRRGGVAVLTVGLAEKVAESLIDVVMRRVPPKIDEXALVRE 550  
 QY 546 LARHVEVTVBENAVAGAGSAINFTFLOAKVLMPCVINGLDPREVEGSGRBEELSLV 605  
 Db 551 LAGSHLLVTLIEBNAMVGAGSANGVGFLLASBGLVEVLQGLPDTYVYERAKSSEMLABCG 610  
 QY 606 LDSKGIATIEBQ 617  
 Db 611 LDAAGIEKAVRQ 622  
 RESULT 8  
 ID AAU38490  
 AAU38490 standard; Protein; 620 AA.  
 AC AAU38490;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Salmonella typhi cellular proliferation protein #381.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;  
 XX antibiotic; antibacterial; drug design.  
 OS Salmonella typhi.  
 XX  
 PN WO200170955-A2.  
 PD 27-SEP-2001.  
 PF 21-MAR-2001; 2001WO-US09180.  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS56349.  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 BS Example 3; Seq ID No 14083; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 620 AA;  
 Query Match 62.6%; Score 2014.5; DA 22; Length 620;  
 Best Local Similarity 61.0%; Pred No. 1.9e-175;  
 Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;  
 QY 7 YELKNHTPADIPALSKDQLOLADEVGYLTHVVISIGSGHFAAGLVEITVALHVF 66  
 Db 8 YPTLALVSTQELRLPRESLPKLCELRVYLVSVSSSGHFAAGLVEITVALHVF 67  
 QY 67 NTPVDQLVWDVQDAVPRKILITRKREKPTITLGGVAFPARDSSEPDAGVGHSSISI 126  
 Db 68 NTPFDQLIWDVQDAVPRKILITRRDKITGRKQGLPFPWRGSEPDVLSVGHSSISI 127  
 QY 127 SAALGMALASOLRGECKMVAIIGDSITGMAVEAMNHAGDVANLVIINDNMSTSP 186  
 Db 128 SAGIGIAVPAKEGKDRITVCVIGGALTAGMAPEAMNHAGDIRPDMVILINDNMSTSP 187  
 QY 187 PVGAMNNTVYLSKKEYSVREBSKALAMPVSWELARTTEHVKACTVPGTLFEEIG 246  
 Db 188 NVGALNNHIALRLSGKLYSSIREGKGVSGVPEIKELKTEEHIKGVVPGTLFEEIG 247  
 QY 247 FNYRSPIDGHVEMVSTLENLKDITGVPFLVVTYKKGAGAPAKDPLAVHGVAPDPT 306  
 Db 248 FNYIGVVDGHDVGMGLISTLNNRDKGQPLATITKGRGEPARKDPTTHAVKPDPS 307  
 QY 307 KDFLEKAPSPHPTTYTEVFGMLCDMAODERLLGITPAMEGSGELVEFSQFPRYFDV 366  
 Db 308 SGCLPEKSS-GGLPGYSKIFGDMLCETAKDSKLVAITPAMEGSGVMEFSRKFPRYFDV 366  
 QY 367 AIAEQHATTLAAGACQCAKPVVAIYSTFLORGPDOLIHDVALQNLMLFALDAGLVGP 426  
 Db 367 AIAEQHATVPAAGIAGVYKPVVAIYSTFLORADVHDVALQNLMLFALDAGLVGA 426  
 QY 427 DGPTTAGAFDYVYRCIPENMLINAPADENECRQMLTGFQHH-GPASYRYRGKGPAGAI 485  
 Db 427 DGQTHQGAFDLSYACIPDMVIMTPSPDENECRQMLPFGYHNDGPYAVRYRGNAQVAL 486  
 QY 486 DPLTLALTEGKAEVRHSGRIALANGSNVTPAVEKQKQATVVMRRVPPDQALVLE 545  
 Db 487 TP-LEKLPITGKLVKRGHEKLAIFGTLMPEAKVVAELAVATVDMKFPVPLDITLIE 545  
 QY 546 LARHVEVTVBENAVAGAGSAINFTFLOAKVLMPCVINGLDPREVEGSGRBEELSLV 605  
 Db 546 MAAQCHALVTLIEBNAMVGAGSANGVGFLLASBGLVEVLQGLPDTYVYERAKSSEMLABCG 605  
 QY 606 LDSKGIATIEBQ 620  
 Db 606 LDAAGIEKAVRQ 620  
 RESULT 9  
 ID AA52832  
 AA52832 standard; Protein; 620 AA.  
 AC AA52832;  
 XX  
 DT 26-JAN-2000 (first entry)  
 XX  
 DE Escherichia coli protein sequence SEQ ID NO:1.  
 XX  
 KW Isoprenoid; microorganism; detection; antibacterial; herbicide;  
 KW heart disease; osteoporosis; haemostasis; cancer; immunopotentialion;  
 KW health food; antifouling coating; farnesyl pyrolonic acid; pyruvic acid;  
 KW 1-deoxy-D-xyulose-5-phosphate; glyceraldehyde-3-phosphate;  
 KW 2-C-methyl-D-erythreitol-4-phosphate.  
 XX  
 OS Escherichia coli.

PN WO953071-A1.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-UP01987.  
 XX  
 PR 14-APR-1998; 98JP-0103101.  
 PR 05-AUG-1998; 98JP-0221910.  
 PR 15-FEB-1999; 99JP-0035739.  
 XX  
 PA (KYO) KYOMA HAKKO KOGYO KK.  
 PI Miyake K, Hashimoto S, Motoyama H, Ozaki A, Seto H, Kuzuyama T,  
 PI Takahashi S;  
 XX WPI: 1999-620434/53.  
 DR N-PSDB; AA233157.  
 XX  
 PT Preparation of recombinant isoprenoid compounds useful for treatment of  
 PT heart diseases, osteoporosis and hemostasis, preventing cancer and  
 PT immunopotentialiation  
 XX  
 PS Claim 2; Page 63-67; 145pp; Japanese.  
 CC The present invention describes the preparation of an isoprenoid  
 CC compound comprising using at least 1 DNA e.g. encoding proteins which  
 CC elevate the efficiency of the synthesis or DNA encoding a farnesyl  
 CC pyrolic producing enzyme. The method of preparation of an isoprenoid  
 CC compound comprises using at least 1 DNA, a vector, cloned cells, their  
 CC derived recombinant DNAs or transformed products in a culture system and  
 CC extracting the isoprenoid accumulated in the medium. The DNA encodes at  
 CC least 1 of the following: (a) a compound for activating or catalysing  
 CC the production of 1-deoxy-D-xylulose-5-phosphate from pyruvic acid and  
 CC glyceraldehyde-3-phosphate; (b) an enzyme producing farnesyl pyrolic  
 CC acid; (c) a protein which elevates the efficiency of synthesis of  
 CC isoprenoid compounds and comprises a 3 or 4 amino acid sequence  
 CC optionally with 1 or more of the amino acids being deleted or  
 CC substituted or an additional amino acid being inserted; (d) a protein  
 CC which activates or catalyses the production of 2-C-methyl-D-erythritol-  
 CC 4-phosphate from 1-deoxy-D-xylulose-5-phosphate; or (e) a protein which  
 CC activates a target compound or reaction and is a string and or hybrid of  
 CC the DNA encoded in (a)-(e). Isoprenoid compounds are useful in drugs  
 CC (e.g. for the treatment of heart diseases, osteoporosis and hemostasis,  
 CC for preventing cancer and as immunopotentializers), health foods and  
 CC antifouling coatings. The isoprenoids also inhibit enzymatic reactions  
 CC on the non-mevalonate pathway and can be used as antibacterials and  
 CC herbicides. The present sequence is used in the exemplification of  
 CC the present invention.  
 CC  
 SO Sequence 620 AA;  
 Query Match 62.2%; Score 2000.5; DB 20; Length 620;  
 Best Local Similarity 60.2%; Pred. No. 3.6e-174;  
 Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;  
 QY 7 YPLIKIHHPADIRASRKOLOQADENVGYLTHTYISGSGHAPAGIGVETVALHYV 66  
 DB 8 YPFIALVSTQERLPLPKESLPDGLRLRYLDSVRSRSGHAPAGIGVETVALHYV 67  
 QY 67 NTFVDLVWDVGHQAYPAKILTRKERMPTIRLTGSVSAFPARDESEYDAFGVSHSTSI 126  
 DB 68 NTFVDLVWDVGHQAYPAKILTRKERMPTIRLTGSVSAFPARDESEYDAFGVSHSTSI 127  
 QY 127 SAALGALISQLRGEKQKVAIIGDSITGMAVEMNHNAGVNNALVITLNDNMSISP 186  
 DB 128 SAIGIGIAVAEKRGKRRITVCVIGDAITAGMAFEMNHNAGIIRPOMLVITLNDNMSISE 187  
 QY 187 PGCAANNVITKLVSSKFSVSESKKALAKMPSVWELARKTEBEHVKGAVIGTIFEBIG 246  
 DB 188 NVCALNNHIALQILSGLYSLRSGKXVSGVPISELUKREBHKKGVGVTIFEBIG 247  
 QY 247 ENYEGTIDGHVEMLVSTENLKOLTGPPFLAVVTYKKGAVPAEADPLATNGVPAFDP 306

DB 248 FNYIGPVGDHVLGLITLLKNMMDLKGQFLHMTTKKRGVEPAEKDPIITHAVKEDPS 307  
 QY 307 KDFPFAAPSEHPITTYVPGRLCDMAAOPBRLLGITPARBEGSLVEFSQKPFNRFPDV 366  
 DB 308 SGCLPKSS-GGLPYSKIPDWMLCETAKDKMKAITIPAREEGSGWVEFRKPFDRFPDV 366  
 QY 367 AIAEOHAVTLAAGACGAKPVVAIYSTFLORGQDILHVALQNLMDLPALDRAGLVGP 426  
 DB 367 AIAEOHAVTLAAGALIGYPIVAIYSTFLORADQVLHVALQKLVLAIDRAGLVGA 426  
 QY 427 DGPTHAAGFDYSYKCEIPNMLINAPADENECROMLTTFQPH-CPASVRYPRGPGAAI 485  
 DB 427 DGQTHOCAFDSYLRCLPEWIMTPSDNECROMLYGHHNDPSAVRYPRGPAVVEL 486  
 QY 486 DPTTALIEGAEVHHGSRITAIAMGSMYTPAVEAKOAGATVNNRFPVGPPOALVLE 545  
 DB 487 TP-LEKLPICKGIVKRGKEKALINFTITLPEAKVABSLNATLYDRFVKYDLEALILE 545  
 QY 546 IARTHDVFTVYENVIAGAGSALNTEFLAQKVLMPYCNIGLDPDPVYEGSREBLILVG 605  
 DB 546 MAASHALIVEEENALINGAGSGVNEVIMHRRKVPVLTIGLPDFFIPOGTQEMERAILG 605  
 QY 606 LDSGILATIEQPCA 620  
 DB 606 LDAAGMEAKIKAWLA 620  
 RESULT 10  
 ID AAY08880 standard; protein; 620 AA.  
 AC AAY08880;  
 XX 13-AUG-1999 (first entry)  
 DE E. coli DXS protein.  
 XX 1-deoxy-D-xylulose-5-phosphate synthase; DXS; modulator; pyruvate;  
 KM glyceraldhyde-3-phosphate; GA3P; 1-deoxy-D-xylulose-5-phosphate; DXP;  
 KM structural analogue; pesticide; antibacterial; herbicide; biosynthesis;  
 KM growth promoter; DXP derivative; isoprenoid; carotenoid; chlorophyll;  
 KM phytol; lutein; sterol; ubiquinone; rubber; taxane.  
 OS Escherichia coli.  
 XX DE29800547-UN.  
 EN 08-APR-1999.  
 PD 16-JAN-1998; 98DB-2000547.  
 PF 28-NOV-1997; 97DB-1052700.  
 PR (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.  
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.  
 PA WPI: 1999-230818/20.  
 DR 1-deoxy-D-xylulose-5-phosphate synthase protein and modulators  
 PT potentially useful as pesticides, antibacterial agents, herbicides,  
 PT etc.  
 XX  
 PS Claim 3; Page 10; 21pp; German.  
 CC This invention describes a 1-deoxy-D-xylulose-5-phosphate synthase (DXS)  
 CC protein or active fragment and a DXS modulator. The DXS modulator is a  
 CC structural analogue of pyruvate, glyceraldehyde-3-phosphate (GA3P) or  
 CC 1-deoxy-D-xylulose-5-phosphate (DXP). Compounds that inhibit DXS activity  
 CC are potentially useful as pesticides, antibacterial agents or herbicides.  
 CC Compounds that stimulate DXS activity are potentially useful as growth  
 CC promoters or for increasing biosynthesis of DXP derivatives, especially  
 CC isoprenoids such as carotenoids, chlorophyll, phytol, lutein, sterols,  
 CC ubiquinone, rubber and taxanes.



QY 606 LDSKGLATIPOFCA 620  
DB 606 LDAAGGEXAKIKAMLA 620

## RESULT 12

AAVS1613  
ID AAVS1613 standard; Protein; 620 AA.

AAVS1613:

26-MAY-2000 (first entry)

E. coli DOXS protein.

DOXS; 1-deoxy-D-xylulose-5-phosphate synthase; HPP; GGPPOR; plant;  
KM p-hydroxyphenylpyruvate dioxygenase; tocopherol; vitamin K; chlorophyll;  
KM geranylgeranyl-pyrophosphate oxidoreductase; carotenoid; transgenic; ss.  
OS Escherichia coli.

MO200008169-A1.

17-FEB-2000.

30-JUL-1999; 99MO-BP05467.

05-AUG-1998; 98DE-1035219.

01-OCT-1998; 98DE-1045216.

01-OCT-1998; 98DE-1045224.

01-OCT-1998; 98DE-1045231.

(SUNG-) SUNGENS GMBH & CO KGAA.

Reindl A, Mejia PL, Palmas JME, Gracia MAC, Ebneih M, Herbers K;

WPI: 2000-195578/17.

N-PSDB; AA288977.

Use of DNA encoding 1-deoxy-D-xylulose-5-phosphate synthase to produce

plants with increased tocopherol, vitamin K, chlorophyll and carotenoid

content

Example 3; Page 76-78; 94pp; German.

This invention describes the novel use of a DNA sequence encoding

1-deoxy-D-xylulose-5-phosphate synthase (DOXS), and optionally

p-hydroxyphenylpyruvate dioxygenase (HPP) and/or geranylgeranyl-

pyrophosphate oxidoreductase (GGPPOR), to produce a plant with increased

tocopherol, vitamin K, chlorophyll and/or carotenoid content. Transgenic

plants containing DOXS DNA coding sequences can be used for production of

plant and bacterial DOXS which also have increased tocopherol, vitamin K,

chlorophyll and/or carotenoid content. The test system can be used to

identify inhibitors of DOXS. This sequence represents the Escherichia

coli DOXS protein described in the method of the invention.

Sequence 620 AA:

Query Match 62.2%; Score 2000.5; DB 21; Length 620;  
Best Local Similarity 60.2%; Pred. No. 3.6e-174;  
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

QY 187 PVGANNVLTVLSSKFEYSVREBSKKALAMPSEVWEIARTEEHVKGMTVPGTLPEELG 246  
DB 188 NVGALNNHIAQLLGGKLYSSLRGEGKVFSGVPPICKLKRTEHKGAVVPGTLPEELG 247  
QY 247 FNYFGPIDGHVEMVSTLBNLKDITGPVPLHVVYKKKGYPAPAKDPLAYHGVAPADPT 306  
DB 248 FNYIGPVVDHVLGLITLKNRRDLKGPQFLHMTXKGRGYEPABKDPITFHAVPKFDPS 307  
QY 307 KDPLPKAAPSPHYTYTEFGMLCDMAAODERLLGITTPARESGGLVEFSQKPEPRYFDV 366  
DB 308 SGCLPKRS-GGLPSYSKIPGDMLCETAAKONKMAITPARESGSOMEFSRKFPPDRYFDV 366  
QY 367 AIAECHANVTLAAGACQACQAKVVAIYSTPLOGDYDOLIHVALNMLPALDPAAGLVGP 426  
DB 367 AIAECHANVTLAAGALIGYKIVAIYSTPLOGADYDOLIHVALNMLPALDPAAGLVGA 426  
QY 427 DQPTHGAPDYSYMRCPNMLIMAPADENECROMLTGPOHH-GRASVRYRGGPGGAI 485  
DB 427 DQPTHGAPDYSYMRCPNMLIMAPADENECROMLTGPOHH-GRASVRYRGGPGGAI 486  
QY 486 DPTLTALFEGKAFVRHHSRIAILAMGSMVTPAVENAGKOLGATVVMRFPVPPDQALVE 545  
DB 487 TP-LEKLPITGKIVRREKALINPGTLMPEAKVAESLNAITVDMRFPVAPLDEALILE 545  
QY 546 LARTHDVFTVEENVVIAAGSAGSAINTELOAKVLPVNCIGLPDRFVQGSREELSLVG 605  
DB 546 MAASHBALVTEENAVINAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGS 605  
QY 606 LDSKGLATIPOFCA 620  
DB 606 LDAAGGEXAKIKAMLA 620

## RESULT 13

AAU34475  
ID AAU34475 standard; Protein; 620 AA.

AAU34475:

14-FEB-2002 (first entry)

E. coli cellular proliferation protein #56.

Antisense; prokaryotic cellular proliferation protein;

KM antibiotic; antibacterial; drug design.

Escherichia coli.

MO200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001MO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haseelbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu HH;

WPI: 2001-611495/70.

N-PSDB; AAS52334.

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 10068; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 620 AA:

Query Match 62.2%; Score 2000.5; DB 22; Length 620;  
Best Local Similarity 60.2%; Pred. No. 3.6e-174;  
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

7 YPLKNIHTPADIRALSDQLOGLADVRGVLTHVTSISGGPAAAGTVELTVAHYVP 66  
8 YPLALVDSSTOELRLPKESLPKLCDELIRYLDSVSSSGHPSGLGVELTVAHYVP 67  
67 NTPVDQVMDVGHQAYPKHILTGKREKMPITRTLGVSAPFARDESEYDAFGVGSSTSI 126  
68 NTPFDQIMDVGHQAYPKHILTGKREKMPITRTLGVSAPFARDESEYDAFGVGSSTSI 127  
127 SAALGMALASQLEBDEKKNVAIIGDGSITGGMAYEAMNHADGVANLVILINDMNSISP 186  
128 SAGIGIAVAAEKGNRRFTVCVIGDGLTAGAFEMNHAGDIRPMDLVILINDMNSISE 187  
187 PVGAMNNVLTLSKSFYSVRESKKALAKMPVWEIARKTEEHKGVIVPGLTFEELG 246  
188 NVGALNNHLAGLISGLYSLSREGKKVSGVPPIKELKTEEHKGVIVPGLTFEELG 247  
247 FNYFGPIDGHVEMLVSTLENLKDLTGPVPLAVVTKGKGYAPAEKDPPLAVGPAFDP 306  
248 FNYIGPVGDHVLGLITLTKNMRDLKGPQFLHMTKGGYEPAEKDPITFAVAPKDP 307  
307 KDFLPAAASHRETTYEVFGRMLCMAAODERLIGTPMREGSGGLVERSGFPNRRYPV 366  
308 SGCLPKRS-GGLPSYSKIGDMLCETPAADNKLMAITTPMREGSGGVERSRKPFRYPV 366  
367 ALAEGAVVTLAAGAOACGAKPVVAISTFLORGYPDLIHVALONDMFLADRLAGLV 426  
367 ALAEGAVVTLAAGAOACGAKPVVAISTFLORGYPDLIHVALONDMFLADRLAGLV 426  
427 DQFTAGAFDYSYWRCLPNMLIAPADNECRQMLTTGFSOH-GRASVYPRGKGPAAI 485  
427 DQFTAGAFDYSYWRCLPNMLIAPADNECRQMLTTGFSOH-GRASVYPRGKGPAAI 485  
486 DPLTALTEIGKAEVRRHSGRIALLANGSVTPRAVEGKQLGATVYNNMRVKKPDALVE 545  
487 TP-LEKLPVIGKGVGRGECALINAGTILMPARAKAESLNTLVDMRVKPLDELLIE 545  
546 LARTHDVFTVEENIAGAGSAINTEFLQAKVLPVCNIGLDFRVEGSEEBELSLVG 605  
546 MAASHBALVTVEENIAGAGSAGVNEVLAHRKPVVLMITGIPDFPIPGGTGBRAEELG 605  
606 LSKGLIATIEQCA 620  
606 LAAAGBEAKIKAWLA 620

RESULT 14

ID AAO21860 standard; Protein; 620 AA.

AAO21860;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 31.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DKS;

decaprenyl diphosphate synthase.

Escherichia coli.

WO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001WO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,

1-deoxyxylulose-5-phosphate synthase activity, useful for the

production of isoprenoids, especially CoQ(10)

disclosure; Fig 6; 246pp; English.

CC The invention relates to methods and materials for the production of  
CC isoprenoids. More particularly the invention provides isolated nucleic  
CC acids, substantially pure polypeptides, host cells, and methods for  
CC producing various isoprenoid compounds. The polypeptides are useful for  
CC the production of isoprenoids, especially CoQ(10). Expressing the pure  
CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
CC increasing production of CoQ(10) in a cell having endogenous DDS  
CC activity. This sequence represents a protein relating to the isoprenoid  
CC production of the invention.

XX Sequence 620 AA:

Query Match 62.2%; Score 2000.5; DB 23; Length 620;  
Best Local Similarity 60.2%; Pred. No. 3.6e-174;  
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

7 YPLKNIHTPADIRALSDQLOGLADVRGVLTHVTSISGGPAAAGTVELTVAHYVP 66  
8 YPLALVDSSTOELRLPKESLPKLCDELIRYLDSVSSSGHPSGLGVELTVAHYVP 67  
67 NTPVDQVMDVGHQAYPKHILTGKREKMPITRTLGVSAPFARDESEYDAFGVGSSTSI 126  
68 NTPFDQIMDVGHQAYPKHILTGKREKMPITRTLGVSAPFARDESEYDAFGVGSSTSI 127  
127 SAALGMALASQLEBDEKKNVAIIGDGSITGGMAYEAMNHADGVANLVILINDMNSISP 186  
128 SAGIGIAVAAEKGNRRFTVCVIGDGLTAGAFEMNHAGDIRPMDLVILINDMNSISE 187  
187 PVGAMNNVLTLSKSFYSVRESKKALAKMPVWEIARKTEEHKGVIVPGLTFEELG 246  
188 NVGALNNHLAGLISGLYSLSREGKKVSGVPPIKELKTEEHKGVIVPGLTFEELG 247  
247 FNYFGPIDGHVEMLVSTLENLKDLTGPVPLAVVTKGKGYAPAEKDPPLAVGPAFDP 306  
248 FNYIGPVGDHVLGLITLTKNMRDLKGPQFLHMTKGGYEPAEKDPITFAVAPKDP 307





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 42.3225 Seconds  
(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947a-6

Perfect score: 3216

Sequence: 1 MKLITDYPILKNIHTPADIR.....LSIVGLDNGKILATTEQFCA 620

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 789580 segs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

|     |  |
|-----|--|
| 1:  | /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*  |
| 2:  | /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*   |
| 3:  | /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*  |
| 4:  | /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*  |
| 5:  | /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*  |
| 6:  | /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.* |
| 7:  | /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*  |
| 8:  | /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*  |
| 9:  | /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*  |
| 10: | /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.* |
| 11: | /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.* |
| 12: | /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*  |
| 13: | /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*  |
| 14: | /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.* |
| 15: | /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.* |
| 16: | /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*  |
| 17: | /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*  |
| 18: | /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | ID                  | Description        |
|------------|--------|-------------|--------|---------------------|--------------------|
| 1          | 3216   | 100.0       | 620    | US-09-934-903-2     | Sequence 2, Appl   |
| 2          | 3216   | 100.0       | 620    | US-09-934-868-62    | Sequence 62, Appl  |
| 3          | 3216   | 100.0       | 620    | US-09-941-947a-6    | Sequence 6, Appl   |
| 4          | 2080   | 64.7        | 628    | US-10-369-493-13998 | Sequence 13998, A  |
| 5          | 2069.5 | 64.4        | 627    | US-10-381-779-15    | Sequence 15, Appl  |
| 6          | 2014.5 | 62.6        | 620    | US-09-815-242-14083 | Sequence 14083, A  |
| 7          | 2004.5 | 62.3        | 619    | US-10-369-493-481   | Sequence 481, A    |
| 8          | 2003.5 | 62.3        | 620    | US-10-369-493-21174 | Sequence 21174, A  |
| 9          | 2000.5 | 62.2        | 620    | US-09-815-242-10068 | Sequence 10068, A  |
| 10         | 2000.5 | 62.2        | 620    | US-10-381-779-31    | Sequence 31, Appl  |
| 11         | 2000.5 | 62.2        | 620    | US-10-369-493-731   | Sequence 731, Appl |
| 12         | 1992.5 | 62.0        | 623    | US-10-369-493-15911 | Sequence 15911, A  |
| 13         | 1992.5 | 62.0        | 623    | US-10-369-493-16281 | Sequence 16281, A  |
| 14         | 1992.5 | 62.0        | 653    | US-10-369-493-15540 | Sequence 15540, A  |
| 15         | 1990.5 | 61.9        | 619    | US-10-369-493-7428  | Sequence 7428, Ap  |

|    |        |      |     |                     |                    |
|----|--------|------|-----|---------------------|--------------------|
| 16 | 1985.5 | 61.7 | 616 | US-10-369-493-4669  | Sequence 4669, Ap  |
| 17 | 1971   | 61.3 | 615 | US-10-369-493-9272  | Sequence 9272, Ap  |
| 18 | 1968.5 | 61.2 | 618 | US-10-369-493-8507  | Sequence 8507, Ap  |
| 19 | 1967   | 61.2 | 613 | US-10-369-493-9367  | Sequence 9367, Ap  |
| 20 | 1951   | 60.7 | 670 | US-10-369-493-17644 | Sequence 17644, A  |
| 21 | 1937.5 | 60.2 | 625 | US-10-381-779-33    | Sequence 33, Appl  |
| 22 | 1776   | 55.2 | 637 | US-10-381-779-32    | Sequence 32, Appl  |
| 23 | 1623   | 50.5 | 624 | US-10-369-493-10188 | Sequence 10188, A  |
| 24 | 1620.5 | 50.4 | 628 | US-10-381-779-18    | Sequence 18, Appl  |
| 25 | 1620.5 | 50.4 | 628 | US-10-369-493-61    | Sequence 61, Appl  |
| 26 | 1555   | 48.4 | 630 | US-10-369-493-20596 | Sequence 20596, A  |
| 27 | 1543.5 | 48.0 | 637 | US-10-369-493-12290 | Sequence 12290, A  |
| 28 | 1532.5 | 47.7 | 641 | US-10-381-779-3     | Sequence 3, Appl   |
| 29 | 1530.5 | 47.6 | 629 | US-10-369-493-10848 | Sequence 10848, A  |
| 30 | 1523.5 | 47.4 | 640 | US-10-369-493-17893 | Sequence 17893, Ap |
| 31 | 1503.5 | 46.8 | 654 | US-10-369-493-14775 | Sequence 14775, A  |
| 32 | 1499.5 | 46.6 | 634 | US-10-369-493-11753 | Sequence 11753, A  |
| 33 | 1499.5 | 46.6 | 634 | US-10-369-493-14237 | Sequence 14237, A  |
| 34 | 1499.5 | 46.6 | 635 | US-10-369-493-14928 | Sequence 14928, A  |
| 35 | 1499.5 | 46.4 | 648 | US-10-381-779-27    | Sequence 27, Appl  |
| 36 | 1491.5 | 46.4 | 720 | US-10-427-590-26    | Sequence 26, Appl  |
| 37 | 1485.5 | 46.2 | 625 | US-10-369-493-18033 | Sequence 18033, A  |
| 38 | 1482   | 46.1 | 625 | US-10-369-493-10317 | Sequence 10317, A  |
| 39 | 1468   | 45.6 | 614 | US-10-369-493-17337 | Sequence 17337, A  |
| 40 | 1459   | 45.4 | 629 | US-10-369-493-20015 | Sequence 20015, A  |
| 41 | 1444   | 44.9 | 635 | US-10-369-493-414   | Sequence 414, Appl |
| 42 | 1444   | 44.9 | 671 | US-10-259-194A-414  | Sequence 29, Appl  |
| 43 | 1442   | 44.8 | 640 | US-10-381-779-29    | Sequence 29, Appl  |
| 44 | 1442   | 44.8 | 640 | US-10-369-493-2645  | Sequence 2645, Ap  |
| 45 | 1439   | 44.7 | 636 | US-10-381-779-28    | Sequence 28, Appl  |

## ALIGNMENTS

RESULT 1  
US-09-934-903-2  
Sequence 2, Application US/09934903  
Patent No. US20020102690A1  
GENERAL INFORMATION:  
APPLICANT: Koffas, Mattheos  
APPLICANT: Odum, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. US20020102690ALton, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Rouviere, Pierre  
APPLICANT: Picataggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
FILE REFERENCE: C11646 US NA  
CURRENT APPLICATION NUMBER: US/09/934, 903  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229,907  
PRIOR FILING DATE: September 1, 2001  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 2  
LENGTH: 620  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
FEATURES:  
OTHER INFORMATION: Amino acid sequences encoded by ORF1

US-09-934-903-2

Query Match 100.0%; Score 3216; DB 10; Length 620;  
Best Local Similarity 100.0%; Pred. No. 1.1e-298;  
Matches 620; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLITDYPILKNIHTPADIRALSDDQOQLADREGLTIVTSISGHPAAGLTVTLV 60  
DB 1 MKLITDYPILKNIHTPADIRALSDDQOQLADREGLTIVTSISGHPAAGLTVTLV 60  
QY 61 ALHVTPTVDQIWMVDVGHQAPHPKILTKRKRMPITRTIGVSAFPARDESRYDAVG 120

```

Db      61  ALHYFNPVDPOLVWDVGHQAYPHKILTGKREMPTRITLGGVSAFPARDESEYDAFVG 120
QY      121  HSSSTISALGMAIALSQRGEDKRMVAITGGGSIITGMAVEAMNHAGVNNMLVTINDN 180
Db      121  HSSSTISALGMAIALSQRGEDKRMVAITGGGSIITGMAVEAMNHAGVNNMLVTINDN 180
QY      181  DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
Db      181  DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
QY      241  LPEELGFNFPGPIDGHVEMLVSTLENLKDLTGVPFLHVVTKKGGYAPAKDPLAHHGV 300
Db      241  LPEELGFNFPGPIDGHVEMLVSTLENLKDLTGVPFLHVVTKKGGYAPAKDPLAHHGV 300
QY      301  PAFDPTKDFLPKAPSPHPTTYTEVFGRMLCDMAODERLLGITPAMREGSGLVESQKFP 360
Db      301  PAFDPTKDFLPKAPSPHPTTYTEVFGRMLCDMAODERLLGITPAMREGSGLVESQKFP 360
QY      361  NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQLIHDVALQNLDMFLALDR 420
Db      361  NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQLIHDVALQNLDMFLALDR 420
QY      421  AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNEBCRMQLTTPQHGGPASYRYPGKG 480
Db      421  AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNEBCRMQLTTPQHGGPASYRYPGKG 480
QY      481  PAAIDPPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
Db      481  PAAIDPPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
QY      541  ALVIELARTHDVPTVTEENVITAGAGSAINTEFLQAKYLVPCNIGLPDRFVEGSSREL 600
Db      541  ALVIELARTHDVPTVTEENVITAGAGSAINTEFLQAKYLVPCNIGLPDRFVEGSSREL 600
QY      601  LSLVGLDSKGLIATIEQFCA 620
Db      601  LSLVGLDSKGLIATIEQFCA 620

```

## RESULT 2

```

US-09-934-868-62
; Sequence 62, Application US/09934868
; Patent No. US2002037190A1
; GENERAL INFORMATION:
; APPLICANT: Kofas, Matcheos
; APPLICANT: Odum, James M
; APPLICANT: Schenzle, Andreas J
; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: CL1596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,668
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 62
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Methylomonas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by DXS
US-09-934-868-62

```

```

Query Match      100.0%; Score 3216; DB 10; Length 620;
Best Local Similarity 100.0%; Pred. No. 1,le-298;
Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  MGLTIDYPLLNKNIHTPADIRALSKDQLQQLADEVAGYLTHTVSISGGHPAAGLGIVELTV 60
Db      1  MGLTIDYPLLNKNIHTPADIRALSKDQLQQLADEVAGYLTHTVSISGGHPAAGLGIVELTV 60
QY      61  ALHYFNPVDPOLVWDVGHQAYPHKILTGKREMPTRITLGGVSAFPARDESEYDAFVG 120

```

```

Db      61  ALHYFNPVDPOLVWDVGHQAYPHKILTGKREMPTRITLGGVSAFPARDESEYDAFVG 120
QY      121  HSSSTISALGMAIALSQRGEDKRMVAITGGGSIITGMAVEAMNHAGVNNMLVTINDN 180
Db      121  HSSSTISALGMAIALSQRGEDKRMVAITGGGSIITGMAVEAMNHAGVNNMLVTINDN 180
QY      181  DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
Db      181  DMSISPPVGAMNNYLTKVLSKFTSSVREBSKKALAKMPSVWELARKTEEHVKMIVPGT 240
QY      241  LPEELGFNFPGPIDGHVEMLVSTLENLKDLTGVPFLHVVTKKGGYAPAKDPLAHHGV 300
Db      241  LPEELGFNFPGPIDGHVEMLVSTLENLKDLTGVPFLHVVTKKGGYAPAKDPLAHHGV 300
QY      301  PAFDPTKDFLPKAPSPHPTTYTEVFGRMLCDMAODERLLGITPAMREGSGLVESQKFP 360
Db      301  PAFDPTKDFLPKAPSPHPTTYTEVFGRMLCDMAODERLLGITPAMREGSGLVESQKFP 360
QY      361  NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQLIHDVALQNLDMFLALDR 420
Db      361  NRYFDVAIAEQHAAVTLAAGACQCAKPVVAIYSTFLQRYDQLIHDVALQNLDMFLALDR 420
QY      421  AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNEBCRMQLTTPQHGGPASYRYPGKG 480
Db      421  AGLVGDPGPTHAGAFDYSYMRCLPNNLIMAPADNEBCRMQLTTPQHGGPASYRYPGKG 480
QY      481  PAAIDPPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
Db      481  PAAIDPPTLTALBIGAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVNNRFPVKPDQ 540
QY      541  ALVIELARTHDVPTVTEENVITAGAGSAINTEFLQAKYLVPCNIGLPDRFVEGSSREL 600
Db      541  ALVIELARTHDVPTVTEENVITAGAGSAINTEFLQAKYLVPCNIGLPDRFVEGSSREL 600
QY      601  LSLVGLDSKGLIATIEQFCA 620
Db      601  LSLVGLDSKGLIATIEQFCA 620

```

## RESULT 3

```

US-09-941-947a-6
; Sequence 6, Application US/09941947A
; Publication No. US2003000528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Kofas, Matcheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odum, J. Martin
; APPLICANT: Picateggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Methylomonas 16a
US-09-941-947a-6

```

```

Query Match      100.0%; Score 3216; DB 11; Length 620;
Best Local Similarity 100.0%; Pred. No. 1,le-298;
Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MKLITDYLKNIHTPADIRALSKDQCOQLADEVRGYLTHTVISGSGHFAAGLGVETLV 60
Db 1 MKLITDYLKNIHTPADIRALSKDQCOQLADEVRGYLTHTVISGSGHFAAGLGVETLV 60
Qy 61 ALHTVFNTPVDQVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGVG 120
Db 61 ALHTVFNTPVDQVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGVG 120
Qy 121 HSSSTISALGMAIASQRGEDKKVVAIIIGDSITGGAAVEAMHAGDVNANLVIYINDN 180
Db 121 HSSSTISALGMAIASQRGEDKKVVAIIIGDSITGGAAVEAMHAGDVNANLVIYINDN 180
Qy 181 DMSISPPVAGANNVLTLYSSKFTSVYSEESKKALAKPSEVWELARKTEEHVKMIVPGT 240
Db 181 DMSISPPVAGANNVLTLYSSKFTSVYSEESKKALAKPSEVWELARKTEEHVKMIVPGT 240
Qy 241 LPEELGPNYFPPIDGHVEMVSTLENKDLTGVPFLHVTYTKKGKGYAPAEKDLAYHGV 300
Db 241 LPEELGPNYFPPIDGHVEMVSTLENKDLTGVPFLHVTYTKKGKGYAPAEKDLAYHGV 300
Qy 301 PAPPTKDFLFXAAPSPHPTYTEVGRMLCDMAADERLIGITPAMEGSGLYVFSQKFP 360
Db 301 PAPPTKDFLFXAAPSPHPTYTEVGRMLCDMAADERLIGITPAMEGSGLYVFSQKFP 360
Qy 361 NRYFDVAIAEGHAYTLAAGACQGAQPVVAIYSTFLORGYDQLIHVALQNLMLFALDR 420
Db 361 NRYFDVAIAEGHAYTLAAGACQGAQPVVAIYSTFLORGYDQLIHVALQNLMLFALDR 420
Qy 421 AGLVGPPGPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFOHGPASVYPRKGG 480
Db 421 AGLVGPPGPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFOHGPASVYPRKGG 480
Qy 481 PGAAIDPTLFLAIEGKEVRRHSGRIAILAMGSMVTPAVEAGKQIGATVVMRFVXKFPDQ 540
Db 481 PGAAIDPTLFLAIEGKEVRRHSGRIAILAMGSMVTPAVEAGKQIGATVVMRFVXKFPDQ 540
Qy 541 ALVLELARTHDVFTVEBNVYAGAGSAINTFLOAQKYLMPVCNIGLDFRVEGGSREEL 600
Db 541 ALVLELARTHDVFTVEBNVYAGAGSAINTFLOAQKYLMPVCNIGLDFRVEGGSREEL 600
Qy 601 LSLVGLDSKGLATITEQFCA 620
Db 601 LSLVGLDSKGLATITEQFCA 620

```

## RESULT 4

```

US-10-369-493-13998
; Sequence 13998, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13998
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13998

```

Query Match 64.7%; Score 2080; DB 12; Length 628;  
 Best Local Similarity 63.2%; Pred. No. 76-190;  
 Matches 387; Conservative 101; Mismatches 120; Indels 4; Gaps 2;

```

Qy 8 PLIKNIHTPADIRALSKDQCOQLADEVRGYLTHTVISGSGHFAAGLGVETLVALHYVN 67
Db 16 PLIDRANTPDGRLRGLAELETTLADELRLTLVGGTGGHFGAGLGVETLITLALHYVD 75
Qy 68 TPVDQVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGVGHSSSTIS 127
Db 76 TPDDRLVMDVGHQAYPHKILTGKREKPTIRTLGGVSAFPARDESEYDAFGVGHSSSTIS 135
Qy 128 AALGMAIASQRLBEDKKVVAIIIGDSITGGAAVEAMHAGDVNANLVIYINDMSSPP 187
Db 136 AALGMAIASQRLBEDKKVVAIIIGDSITGGAAVEAMHAGDVNANLVIYINDMSSRN 195
Qy 188 VGAMNNVLTLYSSKFTSVYSEESKKALAKPSEVWELARKTEEHVKMIVPGTLPEELGF 247
Db 196 VGGESNLTALILSRTYASRBSKQVLSLPGAMELARTTEYAGMLVPTGLFELGWM 255
Qy 248 NYGPIIDGHVEMVSTLENKDLTGVPFLHVTYTKKGKGYAPAEKDLAYHGVPAIDPTK 307
Db 256 NYGPIIDGHVEMVSTLENKDLTGVPFLHVTYTKKGKGYAPAEVDPIDGYHATKLEPLD 315
Qy 308 DFLFXAAPSPHPTYTEVGRMLCDMAADERLIGITPAMEGSGLYVFSQKFPNRYFD 365
Db 316 --APPAAPKAGGPKTSGVGRMLCDMAADERLIGITPAMEGSGLYVFSQKFPNRYFD 373
Qy 366 VALAEGHAYTLAAGACQGAQPVVAIYSTFLORGYDQLIHVALQNLMLFALDRAGLVG 425
Db 374 VALAEGHAYTLAAGACQGAQPVVAIYSTFLORGYDQLIHVALQNLMLFALDRAGLVG 433
Qy 426 PDPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFOHGPASVYPRKGGPRAI 485
Db 434 EDEPTNAGAPDVSVMCIENMLIMAPADNEHCQMLTTGFOHGPASVYPRKGGPRAI 493
Qy 486 DPLFLAIEGKEVRRHSGRIAILAMGSMVTPAVEAGKQIGATVVMRFVXKFPDQALVLE 545
Db 494 EXLEPTEIGKGVRRGSGVNVALLVFGVMAELKXATLDAIVVMRFXKPDALVRE 553
Qy 546 LANTHDVFTVEBNVYAGAGSAINTFLOAQKYLMPVCNIGLDFRVEGGSREELISLVG 605
Db 554 IANSHDLVLTIEENALMGAGVSEFFLARENILKSVLHGLPDVVEHAKPQAQMLAECG 613
Qy 606 LDSKGLATITEQ 617
Db 614 LDEVGIEAAVRE 625

```

## RESULT 5

```

US-10-381-779-15
; Sequence 15, Application US/10381779
; Publication No. US20030219786A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jensen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/0020US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatcSEO for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-381-779-15

```

Query Match 64.4%; Score 2069.5; DB 12; Length 627;  
 Best Local Similarity 63.4%; Pred. No. 7.1e-189;  
 Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;

```

QY 8 PTLKNIHTPADIRALSKDQLOQLADEVRGYLTHTVISISGCHPAAGIGTVELTVALHYVEN 67
DB 16 PDDDASSPAERLRGADLTLADELRKQYLTVYQYSGHAGLGVETLIALHYFD 75
QY 68 TPVDOLVMDVGHQAVPHKILTRKERMPITRTLGVSAPARDESEYDAFGVGHSTIS 127
DB 76 TPDDRLVMDVGHQAVPHKILTRERELMGLTRQKMLAAPPRAESEYDTPGVGHSTIS 135
QY 128 AALGMAIASQLGEGDKKWAIIIGDSITGMAVEAMNAGVYANLVTILNDMSISP 187
DB 136 AALGMAIASQLGEGDKKWAIIIGDSITGMAVEAMNAGVYANLVTILNDMSISP 195
QY 188 VGAMNNYLYTKVLSKRYSSVREESKKALAKMPSVMEIARKEEYHGMIVPGTLFEEIGF 247
DB 196 VGLSLNLYLAKLISSTYSSMRGSKVLSRLPGAMEIARTEEYAKMLVGTTLFEEIGM 255
QY 248 NTFGPIDGHVDMVSTLENKDLTGVPVFLVYTKKGYAPAEKDPYAHGVAPDFPTK 307
DB 256 NTFGPIDGHVDMVSTLENKDLTGVPVFLVYTKKGYAPAEKDPYAHGVAPDFPTK 310
QY 308 DELPKAAP--SPHPTVTVFGRWLCDMAAODERLGITTPAMRBSGLVEFSOKFENRYED 365
DB 311 LEAPGSAPEKKTGEPKSYVFGOMLCDMAAODARLIGITTPAMKSSDLVAFSERIPEYED 370
QY 366 VALAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 425
DB 371 VALAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 430
QY 426 PEGPHTAGFDSYWRCTPNMLIMAPDENECROMLTTFQGHGASRYTRGKPGAAI 485
DB 431 EDGPHTAGFDSYWRCTPNMLIMAPDENECROMLTTFQGHGASRYTRGKPGAAI 490
QY 486 DEPTLALBIGKAEVHRHGSRIATILAMGSMTTPAVEAGKOLGATVYMMRFPDQALVLE 545
DB 491 DEPTLALBIGKAEVHRHGSRIATILAMGSMTTPAVEAGKOLGATVYMMRFPDQALVLE 550
QY 546 LARTHDVFTVEENVLAGAGSAINTFLQAKVIMPCNIGLPDRFVQGSREELSLVIG 605
DB 551 LARTHDVFTVEENVLAGAGSAINTFLQAKVIMPCNIGLPDRFVQGSREELSLVIG 610
QY 606 LDSKGLIATIEQ 617
DB 611 LDAAGIEAKVRQ 622

```

RESULT 6  
US-09-815-242-14083  
Sequence 14083, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14083
; LENGTH: 620
; TYPE: PRF
; ORGANISM: Salmonella typhi
US-09-815-242-14083

```

Query Match 62.6%; Score 2014.5; DB 9; Length 620;  
Best local similarity 61.0%; Pred. No. 13e-183;  
Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;

```

QY 7 YPLKNIHTPADIRALSKDQLOQLADEVRGYLTHTVISISGCHPAAGIGTVELTVALHYV 66
DB 8 YPLTALVDSTQELRLPKESLPKLCDELRRLYLLDSVRSRGHPASGIGTVELTVALHYV 67
QY 67 NTFPVDOLVMDVGHQAVPHKILTRKERMPITRTLGVSAPARDESEYDAFGVGHSTIS 126
DB 68 NTFPVDOLVMDVGHQAVPHKILTRKERMPITRTLGVSAPARDESEYDAFGVGHSTIS 127
QY 127 SAALGMAIASQLGEGDKKWAIIIGDSITGMAVEAMNAGVYANLVTILNDMSISP 186
DB 128 SAALGMAIASQLGEGDKKWAIIIGDSITGMAVEAMNAGVYANLVTILNDMSISP 187
QY 187 PVGAMNNYLYTKVLSKRYSSVREESKKALAKMPSVMEIARKEEYHGMIVPGTLFEEIG 246
DB 188 PVGAMNNYLYTKVLSKRYSSVREESKKALAKMPSVMEIARKEEYHGMIVPGTLFEEIG 247
QY 247 FNYGPIIDGHVDMVSTLENKDLTGVPVFLVYTKKGYAPAEKDPYAHGVAPDFPTK 306
DB 248 FNYGPIIDGHVDMVSTLENKDLTGVPVFLVYTKKGYAPAEKDPYAHGVAPDFPTK 307
QY 307 KDLPKAAPSPHPTVTVFGRWLCDMAAODERLGITTPAMRBSGLVEFSOKFENRYED 366
DB 308 KDLPKAAPSPHPTVTVFGRWLCDMAAODERLGITTPAMRBSGLVEFSOKFENRYED 367
QY 367 ALAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 426
DB 367 ALAEQHAVTLAAGACQCAKPVVAIYSTLQRYGDIHVAALQNDMLPALDRAGLVG 426
QY 427 DGPTHTAGFDSYWRCTPNMLIMAPDENECROMLTTFQGHGASRYTRGKPGAAI 485
DB 427 DGPTHTAGFDSYWRCTPNMLIMAPDENECROMLTTFQGHGASRYTRGKPGAAI 486
QY 486 DEPTLALBIGKAEVHRHGSRIATILAMGSMTTPAVEAGKOLGATVYMMRFPDQALVLE 545
DB 487 DEPTLALBIGKAEVHRHGSRIATILAMGSMTTPAVEAGKOLGATVYMMRFPDQALVLE 545
QY 546 LARTHDVFTVEENVLAGAGSAINTFLQAKVIMPCNIGLPDRFVQGSREELSLVIG 605
DB 546 LARTHDVFTVEENVLAGAGSAINTFLQAKVIMPCNIGLPDRFVQGSREELSLVIG 605
QY 606 LDSKGLIATIEQ 620
DB 606 LDAAGIEAKIKAMLA 620

```

RESULT 7  
US-10-369-493-481  
Sequence 481, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B  
 CURRENT APPLICATION NUMBER: US/10/369,493  
 CURRENT FILING DATE: 2003-02-28  
 PRIOR APPLICATION NUMBER: US 60/360,039  
 PRIOR FILING DATE: 2002-02-21  
 NUMBER OF SEQ ID NOS: 47374  
 SEQ ID NO 481  
 LENGTH: 619  
 TYPE: PRT  
 ORGANISM: Xenorhabdus nematophilus  
 US-10-369-493-481

Query Match 62.3%; Score 2004.5; DB 12; Length 619;  
 Best Local Similarity 61.8%; Pred. No. 1,2e-182;  
 Matches 378; Conservative 102; Mismatches 129; Indels 3; Gaps 3;

QY 7 YPLKNIHNPADIRALSSDOLQOLADEVRGYLHTVTSISGHPAAGLGTVELTVALHYF 66  
 DB 8 YFTLALAEETPEELRLPKDTLPKLCDELROFLNVSRSRGHFAAGLGTVELTVALHYV 67  
 QY 67 NPEVDQVMDVGHQAVPHKILTGKREMPRTITLGVSAPPARDESEYDAFGVGHSTSI 126  
 DB 68 KTFPDLVMDVGHQAVPHKILTGKREMPRTITLGVSAPPARDESEYDAFGVGHSTSI 127  
 QY 127 SAALGMAIASQLRGDKKVAIIGDGSITGGAAYEAMNAGDVANLLVITLNDMSISP 186  
 DB 128 SAGLGMAIAAKHEDKGRKTVCGVIGDGAITAGAFAMNAGDIDPMLVITLNDMSISE 187  
 QY 187 PYGAMNNTLTKYLSKFSVSRESKKAALAKMPSVWEAKTEEHKMGIVPGLTVELG 246  
 DB 188 NGALNNHNLALHLSGKLTTLRSGKVFSGLPPIKELTKTEHLLKMGVAVPGLTVELG 247  
 QY 247 FNYFGPIDGHVEMLVSTLENLKDITGPFVFLHVVTKKGAYAPAEKDPALAHGVAPDPT 306  
 DB 248 FNYIGPVGHVYALVLTQTLKMRBELKGPQLHMTKKGAYAPAEKDPISMAVAPKEDPS 307  
 QY 307 KDFLPKAAAPSPHPTTYEVFGMLCDMAADRELLGITPAMEGSGLVPSQKFPKRYFDV 366  
 DB 308 TGSIPKSSDT-RPTFSKIFGEMLCSEAADKMLAITPAMEGSGMVFSEYEPQYFDV 366  
 QY 367 AIAEONATVTLAAGAACQAKPVVAIYSTFLORGYDOLHDAVALONLMDLFLDRAGLVGP 426  
 DB 367 AIAEONATVTLAAGAACQAKPVVAIYSTFLORGYDOLHDAVALONLMDLFLDRAGLVGP 426  
 QY 427 DGPTHAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRCKGGA 485  
 DB 427 DGPTHAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRCKGGA 486  
 QY 486 DPTLTALLETIGKAEVYRHSGRIAILMGSMVTTPAVENGKQLGATVYMMRFRKFDQALV 545  
 DB 487 QP-LEILPFGKGVIRQGERKIALINFGTLLPVALQAAESLMTVDMRFKPLDKELVLE 545  
 QY 546 LARLTHDVFTEVENYIAGAGSAINYFLQAKVEMPCNIGLPRFVEGSSREELSLV 605  
 DB 546 MASRDLVLTLENNALMGAGSGVNEMLMGTNSVSNLGLPDPVFIPTGTQALHSDLG 605  
 QY 606 LKDSKGLIATIEQ 617  
 DB 606 LKDSKGLIATIEQ 617

## RESULT 8

US-10-369-493-21174  
 Sequence 21174, Application US/10369493  
 Publication No. US20030233675A1  
 GENERAL INFORMATION:  
 APPLICANT: Cao, Yongwei  
 APPLICANT: Hinkle, Gregory J.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Coleman, Barry S.  
 APPLICANT: Chen, Xianfeng  
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B  
 CURRENT APPLICATION NUMBER: US/10/369,493  
 CURRENT FILING DATE: 2003-02-28  
 PRIOR APPLICATION NUMBER: US 60/360,039  
 PRIOR FILING DATE: 2002-02-21  
 NUMBER OF SEQ ID NOS: 47374  
 SEQ ID NO 21174  
 LENGTH: 620  
 TYPE: PRT  
 ORGANISM: Xenorhabdus nematophilus  
 US-10-369-493-21174

Query Match 62.3%; Score 2003.5; DB 12; Length 620;  
 Best Local Similarity 61.1%; Pred. No. 1.5e-182;  
 Matches 377; Conservative 102; Mismatches 131; Indels 7; Gaps 4;

QY 7 YPLKNIHNPADIRALSSDOLQOLADEVRGYLHTVTSISGHPAAGLGTVELTVALHYF 66  
 DB 8 YFTLALAEETPEELRLPKDTLPKLCDELROFLNVSRSRGHFAAGLGTVELTVALHYV 67  
 QY 67 NPEVDQVMDVGHQAVPHKILTGKREMPRTITLGVSAPPARDESEYDAFGVGHSTSI 126  
 DB 68 KTFPDLVMDVGHQAVPHKILTGKREMPRTITLGVSAPPARDESEYDAFGVGHSTSI 127  
 QY 127 SAALGMAIASQLRGDKKVAIIGDGSITGGAAYEAMNAGDVANLLVITLNDMSISP 186  
 DB 128 SAGLGMAIAAKHEDKGRKTVCGVIGDGAITAGAFAMNAGDIDPMLVITLNDMSISE 187  
 QY 187 PYGAMNNTLTKYLSKFSVSRESKKAALAKMPSVWEAKTEEHKMGIVPGLTVELG 246  
 DB 188 NGALNNHNLALHLSGKLTTLRSGKVFSGLPPIKELTKTEHLLKMGVAVPGLTVELG 247  
 QY 247 FNYFGPIDGHVEMLVSTLENLKDITGPFVFLHVVTKKGAYAPAEKDPALAHGVAPDPT 306  
 DB 248 FNYIGPVGHVYALVLTQTLKMRBELKGPQLHMTKKGAYAPAEKDPISMAVAPKEDPS 307  
 QY 307 KDFLPKAAAPSPHPTTYEVFGMLCDMAADRELLGITPAMEGSGLVPSQKFPKRYFDV 366  
 DB 308 TGSIPKSSDT-RPTFSKIFGEMLCSEAADKMLAITPAMEGSGMVFSEYEPQYFDV 366  
 QY 367 AIAEONATVTLAAGAACQAKPVVAIYSTFLORGYDOLHDAVALONLMDLFLDRAGLVGP 426  
 DB 367 AIAEONATVTLAAGAACQAKPVVAIYSTFLORGYDOLHDAVALONLMDLFLDRAGLVGP 426  
 QY 427 DGPTHAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRCKGGA 483  
 DB 427 DGPTHAGAFDYSYMCIPNMLIMAPADENECROMLTGFGH-HGPAVSEYPRCKGGA 484  
 QY 484 AIDPTLTALLETIGKAEVYRHSGRIAILMGSMVTTPAVENGKQLGATVYMMRFRKFDQALV 543  
 DB 485 ELQP-LEILPFGKGVIRQGERKIALINFGTLLPVALQAAESLMTVDMRFKPLDKELVLE 543  
 QY 544 LARLTHDVFTEVENYIAGAGSAINYFLQAKVEMPCNIGLPRFVEGSSREELSLV 603  
 DB 544 LEIASHDMVLTLENNALMGAGSGVNEMLMGTNSVSNLGLPDPVFIPTGTQALHSDLG 603  
 QY 604 VGLDSKGLIATIEQ 620  
 DB 604 VGLDSKGLIATIEQ 620

## RESULT 9

US-09-815-242-10068  
 Sequence 10068, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsem, Kari L.  
 APPLICANT: Zvekind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.

```

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10068
; LENGTH: 620
; TYPE: PRF
; ORGANISM: Escherichia coli
; US-09-815-242-10068

```

```

Query Match      62.2%; Score 2000.5; DB 9; Length 620;
Best Local Similarity 60.2%; Pred. No. 2.8e-182;
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

```

```

7 YLLKNHTPADIRALSKDOLADREGYLTHVTSISGCHPAAGLGTVELTVALHYF 66
8 YPTLAVDSTQGEHLPRKESLPKLCDELRFLDSVSSSGHFAAGLGTVELTVALHYF 67
67 NTPVDQLVMDVGHQAVPHKILTGKREMPRTITLGVSAFPARDESBYDAFGVGHSTSI 126
68 NTPFDQLMDVGHQAVPHKILTGKREMPRTITLGVSAFPARDESBYDAFGVGHSTSI 127
127 SALGMAIASQLEGDEKKNVAIIIGDSITGGMAVEANNAAGVYANLVILINDMSISIP 186
128 SAIGIAVAAEKGGKNNRTVCVIGDAITAGMAFEANNAAGDIRPDMVLINDMSISIP 187
187 PGAMNNYITKVLSSKFSYSSVREBSKKALAMPVSWELAKTBEHVKMIVPGLTPEBIG 246
188 NVGALNNHLAQLSGKLYSLREGKRVPSGVPIKELKRTBEHVKMIVPGLTPEBIG 247
247 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHGVPAFPT 306
248 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHGVPAFPT 307
307 KDFLPKAPSPHPTTYTEVFGKWLCDMAADDERLIGITPAMRSGSLVEPSOKFPNRYPDV 366
308 SCGLPSSS-GGLPSSSKIFGDMWLCETAKDKKALITPAMRSGSLVEPSOKFPNRYPDV 366
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKOLGATVVMRFPKPDQALVLE 545
487 TP-LEKLPKIGKIVRGRKALILNFGTLMPEAKVAESINATLVDMRFPKPDQALVLE 545
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
606 LDSKGLIATIEQFCA 620

```

```

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 620
; TYPE: PRF
; ORGANISM: Escherichia coli
; US-10-381-779-31

```

```

Query Match      62.2%; Score 2000.5; DB 12; Length 620;
Best Local Similarity 60.2%; Pred. No. 2.8e-182;
Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

```

```

7 YLLKNHTPADIRALSKDOLADREGYLTHVTSISGCHPAAGLGTVELTVALHYF 66
8 YPTLAVDSTQGEHLPRKESLPKLCDELRFLDSVSSSGHFAAGLGTVELTVALHYF 67
67 NTPVDQLVMDVGHQAVPHKILTGKREMPRTITLGVSAFPARDESBYDAFGVGHSTSI 126
68 NTPFDQLMDVGHQAVPHKILTGKREMPRTITLGVSAFPARDESBYDAFGVGHSTSI 127
127 SALGMAIASQLEGDEKKNVAIIIGDSITGGMAVEANNAAGVYANLVILINDMSISIP 186
128 SAIGIAVAAEKGGKNNRTVCVIGDAITAGMAFEANNAAGDIRPDMVLINDMSISIP 187
187 PGAMNNYITKVLSSKFSYSSVREBSKKALAMPVSWELAKTBEHVKMIVPGLTPEBIG 246
188 NVGALNNHLAQLSGKLYSLREGKRVPSGVPIKELKRTBEHVKMIVPGLTPEBIG 247
247 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHGVPAFPT 306
248 FNTFPGIDGHVEMLVSTLENKDLTGVPFLHVVTKKGVAPAEKDPPLAHGVPAFPT 307
307 KDFLPKAPSPHPTTYTEVFGKWLCDMAADDERLIGITPAMRSGSLVEPSOKFPNRYPDV 366
308 SCGLPSSS-GGLPSSSKIFGDMWLCETAKDKKALITPAMRSGSLVEPSOKFPNRYPDV 366
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
367 AIAEOHVAITLAAGAACQAGKPVVAIYSTFLQAGYDOLHVAQLQNTDMLFALDRAGLVGP 426
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
427 DGPTHAGAFDYVYMCITPNMLIMAPADENECROMLITTFQGH-GPASYRVPKRGKGAAL 485
486 DPTLALBGAERHHSRIAILAMGSMVTPAVEAGKOLGATVVMRFPKPDQALVLE 545
487 TP-LEKLPKIGKIVRGRKALILNFGTLMPEAKVAESINATLVDMRFPKPDQALVLE 545
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
546 LAKTHDVYVYVEENYIAGAGSAINTPLOQKVTLMPCVNIIGLDFRVEQGSREELSLVG 605
606 LDSKGLIATIEQFCA 620

```



Db 606 LDAGMEAKIKAMLA 620

RESULT 11

US-10-369-493-731  
 : Sequence 731, Application US/10369493  
 : Publication No. US20030233675A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cao, Yongwei  
 : APPLICANT: Hinkle, Gregory J.  
 : APPLICANT: Slater, Steven C.  
 : APPLICANT: Goldman, Barry S.  
 : APPLICANT: Chen, Xianfeng  
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 : FILE REFERENCE: 38-10(52052)B  
 : CURRENT APPLICATION NUMBER: US/10/369,493  
 : PRIOR FILING DATE: 2003-02-28  
 : PRIOR FILING DATE: 2002-02-21  
 : NUMBER OF SEQ ID NOS: 47374  
 : SEQ ID NO 731  
 : LENGTH: 620  
 : TYPE: PRT  
 : ORGANISM: Escherichia coli  
 : US-10-369-493-731

Query Match 62.2% Score 2000.5; DB 12; Length 620;

Best Local Similarity 60.2%; Pred. No. 2.8e-182;

Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

7 YPLKNIHTPADIRALSKDLOQLADENVGYLTHVTSISGHPAAGLGTVELTVAHYF 66  
 8 YPLALVDSYQELRLPKESLPCDELARVLDVSRSRGHFASTGLVETLVAHYF 67  
 67 NPVVDQVWDVGHQYPHKILTGKREKPTITLGGVSAFPAPDSEYDAPGVGSSSI 126  
 68 NPVVDQVWDVGHQYPHKILTGKREKPTITLGGVSAFPAPDSEYDAPGVGSSSI 127  
 127 SAALGKAIASQLRGDKKMAVIGDGSITGMAVYEMNAGVNVNLVITLNDNMSTSP 186  
 128 SAGIGVAVLAEGKGRKRTVCVIGDAITAGHAFEMNPAQDIPDMVLINDNMSTSP 187  
 187 PVGAMNNTLTKVLSKFFYSVRESKALAKMPSVWELARKTEHYKAVIYDGLTVEELG 246  
 188 NVGALNNHLLQLLSGLYSLRGGKVSVPPIKELKREHKKAVVYGTILFEELG 247  
 247 FNVFGSIDHDVEMVSTLENKDLTGPFVLAHVYTKGKGYPAPDPLATYGPAPDPT 306  
 248 FNVFGSIDHDVEMVSTLENKDLTGPFVLAHVYTKGKGYPAPDPLATYGPAPDPT 307  
 307 KDFLPAASPHPTTYEVFGMWLCDMAADBERLGTTPMREGSGLVESQKFPNRYFV 366  
 308 SGLTPKSS-CGLPSYKIRIGDMLCETRAKDNLMTTPMREGSGLVESQKFPNRYFV 366  
 367 AIAEGAVTLAAGQACQAKPVVAIYSTFLRGYDQLIHVALQNLMLPALDRAEVLGP 426  
 427 DQPTVAGADVSVMRCIPMLIMAPADENECROMLTGFGOH-GRAPSVYPPGKRGAAI 485  
 427 DQPTVAGADVSVMRCIPMLIMAPADENECROMLTGFGOH-GRAPSVYPPGKRGAAI 486  
 486 DPTLTALEIGKAEVHRHSRIAILANGSVTPAVENAGKQLGATVVMNRPVFKPDQVLE 545  
 487 TP-LEKLPICXGIVKRRGKELIINFGTLMPEAKVAESIANIVDMRVKRLDLALILE 545  
 546 LARHTVPTVTEBNTVAGAGSALNTFLQAKTLMVNCIGPDRFVREGSSEELSLVNG 605  
 546 MAASHALVTVEENALMGAGSGVNEVLMHRRPVVAINIGLPDEFIPGTOEBKRAELG 605  
 606 LNSKGLATIEGFC 620

Db 606 LDAGMEAKIKAMLA 620

RESULT 12

US-10-369-15911  
 : Sequence 15911, Application US/10369493  
 : Publication No. US20030233675A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cao, Yongwei  
 : APPLICANT: Hinkle, Gregory J.  
 : APPLICANT: Slater, Steven C.  
 : APPLICANT: Goldman, Barry S.  
 : APPLICANT: Chen, Xianfeng  
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 : FILE REFERENCE: 38-10(52052)B  
 : CURRENT APPLICATION NUMBER: US/10/369,493  
 : PRIOR FILING DATE: 2003-02-28  
 : PRIOR FILING DATE: 2002-02-21  
 : NUMBER OF SEQ ID NOS: 47374  
 : SEQ ID NO 15911  
 : LENGTH: 623  
 : TYPE: PRT  
 : ORGANISM: Xanthomonas campestris  
 : US-10-369-493-15911

Query Match 62.0% Score 1992.5; DB 12; Length 623;

Best Local Similarity 61.4%; Pred. No. 1.7e-181;

Matches 378; Conservative 91; Mismatches 142; Indels 5; Gaps 3;

4 TTDVYLNKNIHTPADIRALSKDLOQLADENVGYLTHVTSISGHPAAGLGTVELTVAH 63  
 5 STRPRLSRITQTPDDLRFEEDLTAVADSLRAVLISVGSAGHPAAGLGTVELTVAH 64  
 64 YVNPVVDQVWDVGHQYPHKILTGKREKPTITLGGVSAFPAPDSEYDAPGVGSSSI 123  
 65 YVNPVVDQVWDVGHQYPHKILTGKREKPTITLGGVSAFPAPDSEYDAPGVGSSSI 124  
 124 TSISALGMAIASQLRGDKKMAVIGDGSITGMAVYEMNAGVNVNLVITLNDNMSTSP 181  
 125 TSISALGMAIASQLRGDKKMAVIGDGSITGMAVYEMNAGVNVNLVITLNDNMSTSP 184  
 182 MSISPPVGAANNNTLTKVLSKFFYSVRESKALAKMPSVWELARKTEHYKAVIYDGLTVEELG 239  
 185 MSISPPVGAANNNTLTKVLSKFFYSVRESKALAKMPSVWELARKTEHYKAVIYDGLTVEELG 244  
 240 TLFEELGNTYGPIDHDVEMVSTLENKDLTGPFVLAHVYTKGKGYPAPDPLATYGPAPDPT 299  
 245 TLFEELGNTYGPIDHDVEMVSTLENKDLTGPFVLAHVYTKGKGYPAPDPLATYGPAPDPT 304  
 300 VPAFPTDQFPAKAPSPHTTYEVFGMWLCDMAADBERLGTTPMREGSGLVESQKFPNRYFV 359  
 305 VGPDPSPKGLVAKAG-AKKPTTYDVSQVCDMAAADPMLVITTPMREGSGLVESQKFPNRYFV 363  
 360 PNRYFDVAIAEGAVTLAAGQACQAKPVVAIYSTFLRGYDQLIHVALQNLMLPALDRAEVLGP 419  
 364 PNRYFDVAIAEGAVTLAAGQACQAKPVVAIYSTFLRGYDQLIHVALQNLMLPALDRAEVLGP 423  
 420 RAGVGPSPGPHHAADVSVMRCIPMLIMAPADENECROMLTGFGOH-GRAPSVYPPGKRGAAI 479  
 424 RAGVGPSPGPHHAADVSVMRCIPMLIMAPADENECROMLTGFGOH-GRAPSVYPPGKRGAAI 483  
 480 GPGALDPTLTALEIGKAEVHRHSRIAILANGSVTPAVENAGKQLGATVVMNRPVFKPDQVLE 539  
 484 GPGALDPTLTALEIGKAEVHRHSRIAILANGSVTPAVENAGKQLGATVVMNRPVFKPDQVLE 543  
 540 QALVLELARTHDVYVTEBNTVAGAGSALNTFLQAKTLMVNCIGPDRFVREGSSEELSLVNG 599  
 544 KAMLELAKCHRAVSVLEEDNVVAGAGSVELLNAESVLMPEMLGLPDSFORHASEED 603  
 600 LLSLVGLDSKGLIATI 615



Db 566 KXMLELACHEAFVIEDNVAGAGSVSEILANESVLPMLHGLDPSFGHNSRED 625  
QY 600 LSLVGLDSKGLIATI 615  
Db 626 LLAEGIDQAGIRPAV 641

RESULT 15  
US-10-369-493-7428

/ Sequence 7428, Application US/10369493  
/ Publication No. US2003023675A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Cao, Yongwei  
/ APPLICANT: Hinkle, Gregory J.  
/ APPLICANT: Slater, Steven C.  
/ APPLICANT: Goldman, Barry S.  
/ APPLICANT: Chen, Xianfeng  
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
/ FILE REFERENCE: 38-10(52052)B  
/ CURRENT FILING DATE: 2003-02-28  
/ PRIOR FILING DATE: 2002-02-21  
/ NUMBER OF SEQ ID NOS: 47374  
/ SEQ ID NO 7428  
/ LENGTH: 619  
/ TYPE: PR1  
/ ORGANISM: Burkholderia cepacia  
US-10-369-493-7428

Query Match 61.9%; Score 1990.5; DB 12; Length 619;  
Best Local Similarity 61.7%; Pred. No. 2.6e-181;  
Matches 381; Conservative 89; Mismatches 141; Indels 7; Gaps 3;

QY 4 TTDYPLKXNHTPADRALSKOQLOADVRGYLTHWVSISGHPAAGLGVETVALH 63  
Db 3 TTMYDLKTIIDPAALRRLLDRQLADELRAFLVDSVOTGHLSSNLGVETLIALH 62  
QY 64 YVNTFVDOGLVMGVHQAAYPHKILTGKRMPTITLGGVSAFAPARDESEYDAFGVSHS 123  
Db 63 YVFDTHDRIVMVGQHTYPHKILTGKRDQMTLRQLGGISGFPKDESEYDFGTASHS 122  
QY 124 TSISAALGMAISQLEGEDEKQVAILIGDSITGSAVEANAG- DYNANLLVILNDNM 182  
Db 123 TSISALGMAVASKLGDNRMGIAVIGDAMTAGAFBAMNNAQVEDVPLVILNDNM 182  
QY 183 STSPPGAMNNTLTKLSSTFYSSVAREBSKKALAKMPSWELAKTBEYVKGMIYPTLF 242  
Db 183 STSPPGALNRHLARLSGPFYAARAQVERVLRVAPMLDLARKLEBAKMIYPATLF 242  
QY 243 EELGFNYFGPIDGHVEMLVSTLENLKDLTGPVFLHVETKKGVAPEKDPPLAYHGA 302  
Db 243 EERGFNYIGPIDGHDLDSLITLQNKELRGPFGLHVETKKGQYKLAADPVLYHGRK 302  
QY 303 FDDTKDPLKAAAPSPHTTTEVFGRMLCMAAODERLIGITPAMRSGSLVFSQFPR 362  
Db 303 FPAEGIKPAATPS-KKTYTQVGEWLCDAAELDARVIGITPAMRSGSVMVEFEKFPDR 361  
QY 363 YEDVALAEQHAATLAAQACOGAKPVVALYSTFLQGYDQLIHVALQMLDMLFALDRAG 422  
Db 362 YEDVGAIEQHAATFAGGLAABGKPVVALYSTFLQADQLIHVALQMLPVVFAIDRAG 421  
QY 423 LVGPDGPTAGAFDYSYMRCPNMLIMAPADENECROMLTGFOHNGSAVYPRGKPG 482  
Db 422 LVGADGATAGAYDLFLRCIPMTVMASDENECROMLYTALQOPPAVAYPRGAGTG 481  
QY 483 AALDPTLTALGKAEVRHH-----GSRITALLAMGSMTPAVEAGKQIGATVMMRFYCP 537  
Db 482 VAITIKOMTALPLGSKGIRRETSQPAKRIALILAFGTWVAPSLAAEQLDATVANNMFYCP 541  
QY 538 FQALVLELARTHDEVTEVTEENVIAAGASAINTPLOAKQVIMPVCNIGLPDRFVEQGR 597

Db 542 LDADIVROLAETHDAIVTEEGCVWGAAGSACVEALLASGTRPVLQGLPDRFIDHDP 601  
QY 598 EELSLVGLDSKGLIATI 615  
Db 602 AKLLAAGIDAVGITKSI 619

Search completed: January 29, 2004, 16:21:14  
Job time : 44.3225 secs



QY 366 VALAQAHTVLAAGACGACKEPVVALYSTFLQRGYDQLIHVALQNLMLPALDAGLVG 425  
DB 383 VALAQAHTVLAAGACGACKEPVVALYSTFLQPADQLIHVALVDELVLFDIDAGLVG 442  
QY 426 PDGPTAGAGFDSYKRCIPNMLIMAPADENECROMLTGFGHNGASVAPPGKPGCAI 485  
DB 443 EDGPTAGAGFDSYKRCIPNMLVMPSPDEDLKLLTGYLFDGPAARVAPPGSGPNHI 502  
QY 486 DDTLALBEGKAEVPHHSGRIAILANGSWTTPAVEAKGOLATVYNNRPVAFEDQALVLE 545  
DB 503 DDLOPVEIGKGVRRRGVVALVFGVQLAEMVVASLDATVYDMFVPLDVALVRE 562  
QY 546 IARTDVFVTEBENYIAGAGSAINTPLQAKVLEPVNCIGLPPDFVQSGSEELSLVVG 605  
DB 563 IAGSHLVLTIENNAVMGASGAVGEPLASBGLFVPLDGLDPTVYEHAPSEMLACG 622  
QY 606 LDKSGILATIEQ 617  
DB 623 LDAAGIEKAVRQ 634

## RESULT 2

US-09-328-352-6210  
Sequence 6210, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GIG99-03PA  
CURRENT APPLICATION NUMBER: US/09/328-352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 6210  
LENGTH: 648  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-6210

Query Match 54.4%; Score 1751; DB 4; Length 648;  
Best Local Similarity 55.1%; Pred. No. 4,6e-167;  
Matches 342; Conservative 94; Mismatches 165; Indels 20; Gaps 6;

QY 8 PLKNIHTPADIRALSQDLQQLADVRGYLTHTVTSISGHEFAAGLGVETLVALHYVN 67  
DB 28 PLIDLDIDHPOQLRQLEHSQLQVADDELROYILYAAGSGGHEGAMLVGVELTVALHYEN 87  
QY 68 TPVDQLVMDVGHQAYPHKILTGKREMPRTITLGVSAFPAPADESEYDAFGVGHSTIS 127  
DB 88 TPNDRLVMDVGHQAYPHKILTGKREMPRTITLGVSAFPAPADESEYDAFGVGHSTIS 147  
QY 128 AALGMAIASQLRGEDEKQVAILIGDSITGMAVEAMNHAGDVNANLVLINDNDKSI 187  
DB 148 AGLGMSLARQYKDPCEVVCIVGDGAMTGVAFEMANNADVAHADLIVLINDNDKSI 207  
QY 188 VGNANNTLVKVASSTFYSSVRESEKALAKMPSVELAKTEEHKGMIVERTLFEELGF 247  
DB 208 TGGPFAHGLAIWEKGLVAVNNEHGEAYIQPHK-W--TYSRLHOSATDAADNLFKALGF 264  
QY 248 NYFGPIDGHDVEMLVSTLENLKDILGPFVFLHVTYKKGAVAPABKDPALYHGVAPADPTK 307  
DB 265 DVFPGPDGDVQVQVFNALKKRGKPRLVHYTTKKGGFAPABADPTTYHALIGINMA- 323  
QY 308 DFLPKAASPHPTTYTEVFRMLCDMAADDELLGITTPAMRESSGLVRSQKFPNYPVYA 367  
DB 324 ---SGGKTP-PRYSIVFGEMLCDEAADOERLLATTPACBSGMYKAKOPPPHFPVYA 378  
QY 368 IAEQHAHTVLAAGACGACKEPVVALYSTFLQRGYDQLIHVALQNLMLPALDAGLVG 427  
DB 379 IAEQHAHTVLAAGACGACKEPVVALYSTFLQRGYDQLIHVALQNLMLPALDAGLVG 438  
QY 428 GETHAGAFYSYKRCIPNMLIMAPADENECROMLTGFGHNGASVAPPGKPGCAI 487

DB 439 GPTAGAGVDAVYMTVEENMIMAPADENECROMLHTAYVANGPAARVAPPGAGVGEI 498  
QY 488 TLPLALBEGKAEV----RHGSRILAIANGSWTTPAVEAKGQ-----LGATVNNRPVK 536  
DB 499 EMVLELEGKAEIYAEIRANDDEQITVLAFGSRVVALAEAEQFAKQADVSCVNNRFPVK 558  
QY 537 PFDQALVELARTDVFVTEBENYIAGAGSAINTPLQAKVLEPVNCIGLPPDFVQSGS 596  
DB 559 PLDEQIRDLAEHTHELVTEBENYIAGAGSAINTPLQAKVLEPVNCIGLPPDFVQSGS 618  
QY 597 REELSLVGLDSKGIATIEQ 617  
DB 619 HNKWLQDQGDAGKILNSTER 639

## RESULT 3

US-09-857-556A-26  
Sequence 26, Application US/09857556A  
Patent No. 6558915  
GENERAL INFORMATION:  
APPLICANT: Rebecca E. Cahoon  
APPLICANT: Sean U. Coughlan  
APPLICANT: Yong Tao  
APPLICANT: Zude Weng  
APPLICANT: Mark E. Williams  
TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-phosphate Synthase  
FILE REFERENCE: BR1290  
CURRENT APPLICATION NUMBER: US/09/857,556A  
CURRENT FILING DATE: 2001-06-04  
PRIOR APPLICATION NUMBER: 60/110,779  
PRIOR FILING DATE: 1998-12-03  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 26  
LENGTH: 720  
TYPE: PRT  
ORGANISM: Oryza sativa  
US-09-857-556A-26

Query Match 46.2%; Score 1485.5; DB 4; Length 720;  
Best Local Similarity 49.2%; Pred. No. 2,8e-140;  
Matches 314; Conservative 102; Mismatches 185; Indels 37; Gaps 13;

QY 8 PLKNIHTPADIRALSQDLQQLADVRGYLTHTVTSISGHEFAAGLGVETLVALHYVN 67  
DB 71 PLIDLVVPIHMKLSLKEHSQLQVADDELROYILYAAGSGGHEGAMLVGVELTVALHYVN 130  
QY 68 TPVDQLVMDVGHQAYPHKILTGKREMPRTITLGVSAFPAPADESEYDAFGVGHSTIS 127  
DB 131 TPQKILMDVGHQAYPHKILTGKREMPRTITLGVSAFPAPADESEYDAFGVGHSTIS 190  
QY 128 AALGMAIASQLRGEDEKQVAILIGDSITGMAVEAMNHAGDVNANLVLINDNDKSI 184  
DB 191 AALGMAVGRDLKGGKNNVAVTIGDAMTAGCAVEANNAGLDSMDIIVLINDNDKSI 250  
QY 185 -----SPGKAMNTLVKVASSTFYSSVRESEKALAKMPSVELAKTEEHKGMIVERTLFEELGF 236  
DB 251 ATLDPAPPVGASLSALSKQSSPFLAEVAKGVTKQIGSGSHAEAAKVADEPARAKNIS 310  
QY 237 -VPGTLEELGFNYFGPIDGHDVEMLVSTLENLKD-LTGFVFLHVTYKKGAVAPABK 293  
DB 311 GSGSTLFEELGLVYTGVDGHNIDDLITLREYKSTTGTSPVLIHVTYKKGAVAPABK 370  
QY 294 PLAHGVPADPT--KQFLPKAASPHPTTYTEVFRMLCDMAADDELLGITTPAMRESSG 351  
DB 371 ADKRVHGAFTDPATKQF---KSPAKTLSTYNNFAALINEABQDNVVALHAAMGGTG 427  
QY 352 LVRSQKFPNYPVYDAEQAHTVLAAGACGACKEPVVALYSTFLQRGYDQLIHVALQNL 411  
DB 428 LMYLRFPKRCPEVGLAEQAHTVLAAGACGACKEPVVALYSTFLQRGYDQVAVDVLQK 487  
QY 412 LDMFLPADRGLVSPDGPPTAGAFDYSYKRCIPNMLIMAPADENE-CROMLTGFGHNG 470

Db 488 LPRFAMDRAGLVGADGPTHGCAFDVTYMACLPNNVMAPSDEAEICMVAATAAIDRRP 547  
Qy 471 ASVRYPRGKGPAAIDPTL--TLEIGKAEVRHHSRIAILMGSVTAVRAGK----- 523  
Db 548 SCFRYRGRGICVPLPPTKGVPLVGRVLLGSRVALLGSGAVQYCLAAASIVRHH 607  
Qy 524 QLCATVYNNRFPVPEPDQALVLEIARCTHDVFTVEENVVAGGASAINTELQAKQYL---- 579  
Db 608 GLKATVADARFCPLDQTLIRRLASSHEVLTVEBSGI-GGRGSHVAGFMALDGLDKL 666  
Qy 580 --MPVGNIGLDPDFRFGSGREELSLVGLDSKILATI 615  
Db 667 KMRPLV--LPDRYIDHGSPPAQALASAGLTPSHIAATV 701

## RESULT 4

US-09-857-556A-34  
; Sequence 34, Application US/09857556A  
; Patent No. 6558915  
; GENERAL INFORMATION:  
; APPLICANT: Rebecca B. Cahoon  
; APPLICANT: Sean J. Coughlan  
; APPLICANT: Yong Tao  
; APPLICANT: Zude Weng  
; APPLICANT: Mark E. Williams  
; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase  
; FILE REFERENCE: B81290  
; CURRENT APPLICATION NUMBER: US/09/857,556A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 60/110,779  
; PRIOR FILING DATE: 1998-12-03  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 34  
; LENGTH: 594  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-09-857-556A-34

Query Match 44.4%; Score 1428; DB 4; Length 594;  
Best Local Similarity 50.7%; Pred. No. 1,2e-134;  
Matches 298; Conservative 93; Mismatches 169; Indels 28; Gaps 11;

Qy 16 PADIRALSDQQLADENVGRLTHTVSISGCHFAAGLGTVELTVALHVTPTVDQLYM 75  
Db 3 PIMKRLSKELQQLADELRSDVIFVSKTGHLSGLGVBLTVALHVTPTVDQKLM 62  
Qy 76 DVGHQAVPHKILTRKERMTIRTLGVSAPFAPRDESYDAFGVSHSTISIAALGMATA 135  
Db 63 DVGHQAVPHKILTRGRDKMTIRTLGVSAPFAPRDESYDAFGVSHSTISIAALGMATA 122  
Qy 136 SOLRGEDKQWALIGDSITGGMAYEAMNHAGDVANLVIINDN-DMSI-----SP 186  
Db 123 RLKGGKNNVAVIIGGAMTAGQAYEAMNHAGDVIINDN-QVSLPATYLDGPAP 182  
Qy 187 PVGAMNNYLTKLVSKEYSVREESKKALAKM-PSVWEIARKTEBEHYKMI--VPQTLE 243  
Db 183 PVGALLSALSKQSSPLRLREVRKGVTKQIGGSVHEIAAAYDEVARQMIGSSSTLE 242  
Qy 244 ELGENTYIGDGHVEMLVSTLEND--LTGPVPLHVTYKKGKGYAPARCKPLAYHGP 301  
Db 243 ELGLVYIGVDDGNIDDLITLREVKSTYKTPGVLIHVTEKRGYPVABRAADKHHGA 302  
Qy 302 ASDPT--KDFLRKAPSPHPTTEVGRMLCDMAADERLLGTPAMRSGSLVFSQKF 359  
Db 303 KEDPATGKF--KSPAKTLSTYNNFAELIABEQDNVVAIHAAMGGSTGLNYFLRRF 359  
Qy 360 PRRYEDVAIAEGHVAITLAQAQACQAKPVVAISTFLQGYDQLIHDAVLAQMLDMLFALD 419  
Db 360 PRRYEDVAIAEGHVAITLAQAQACQAKPVVAISTFLQGYDQVHDVLDQLPVRFAMD 419  
Qy 420 RAGLVGDPGPTHGCAFDVSYRNCIPNMLMAPADENE-CRQMLTTFQGHGASVRYPRG 478

Db 420 RAGLVGADGPTHGCAFDVTYMACLPNNVMAPSDEAEICMVAATAAIDRRPSCFRYPRG 479  
Qy 479 KGRGAIDPTL--TLEIGKAEVRHHSRIAILMGSVTAVRAGK-----QLCATVYN 531  
Db 480 NGIGPLPPTKGVPLVGRVLLGSRVALLGSGAVQYCLAAASIVRHHGLKVTAD 539  
Qy 532 MRFVPEPDQALVLEIARCTHDVFTVEENVVAGGASAINTELQAKQYL 579  
Db 540 ARFCPLDQTLIRRLASSHEVLTVEBSGI-GGRGSHVAGFMALDGLDGL 586

## RESULT 5

US-09-857-556A-12  
; Sequence 12, Application US/09857556A  
; Patent No. 6558915  
; GENERAL INFORMATION:  
; APPLICANT: Rebecca B. Cahoon  
; APPLICANT: Sean J. Coughlan  
; APPLICANT: Yong Tao  
; APPLICANT: Zude Weng  
; APPLICANT: Mark E. Williams  
; TITLE OF INVENTION: Plant 1-Deoxy-Xylulose 5-Phosphate Synthase  
; FILE REFERENCE: B81290  
; CURRENT APPLICATION NUMBER: US/09/857,556A  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 60/110,779  
; PRIOR FILING DATE: 1998-12-03  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 12  
; LENGTH: 708  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-857-556A-12

Query Match 44.2%; Score 1420.5; DB 4; Length 708;  
Best Local Similarity 46.9%; Pred. No. 9.5e-134;  
Matches 300; Conservative 113; Mismatches 187; Indels 39; Gaps 14;

Qy 8 PLKNHHPADIRALSKDQLQQLADENVGRLTHTVSISGCHFAAGLGTVELTVALHVTPTVDQLYM 67  
Db 63 PLDQVYVNPIMKRLSKELQQLADELRSDVIFVSKTGHLSGLGVBLTVALHVTPTVDQKLM 122  
Qy 68 TPVDQVMDVGHQAVPHKILTRKERMTIRTLGVSAPFAPRDESYDAFGVSHSTISIS 127  
Db 123 APDKILMDVGHQAVPHKILTRGRDKMTIRTLGVSAPFAPRDESYDAFGVSHSTISIS 182  
Qy 128 AALGMAIASOLRGEDKQWALIGDSITGGMAYEAMNHAGDVANLVIINDN-DMSIS- 185  
Db 183 AGLGMAVGRDLKGRKNNVAVIIGGAMTAGQAYEAMNHAGDVIINDN-QVSLPATYLDGPAP 242  
Qy 186 -----PVGAMNNYLTKLVSKEYSVREESKKALAKMPS-VWEIARKTEBEHYKMI- 236  
Db 243 ATLDPPIPVGALLSALSRLOSNPLRLREVRKGVTKQIGGSVHEIAAAYDEVARQMIGSSSTLE 302  
Qy 237 -VPGTLEFELGPNYRGPIDGHVEMLVSTLEND--LTGPVPLHVTYKKGKGYAPARCKPLAYHGP 293  
Db 303 GSGSSTLEFELGPNYRGPIDGHVEMLVSTLEND--LTGPVPLHVTYKKGKGYAPARCKPLAYHGP 362  
Qy 294 PLAYHGVAPADPT--KDFLRKAPSPHPTTEVGRMLCDMAADERLLGTPAMRSGSLVFSQKF 351  
Db 363 ADKHGVTGKPDPRPGKQKSKATQSYTYY--FAELIABEQDNVVAIHAAMGGSTGLNYFLRRF 419  
Qy 352 LVERFSQKPNRYEYVAIAEGHVAITLAQAQACQAKPVVAISTFLQGYDQLIHDAVLAQMLDMLFALD 411  
Db 420 MTLFRRRPPTECFVGVIAEGHVAITLAQAQACQAKPVVAISTFLQGYDQVHDVLDQLPVRFAMD 479  
Qy 412 LDMFLALDRAGLVGDPGPTHGCAFDVSYRNCIPNMLMAPADENE-CRQMLTTFQGHGASVRYPRG 470  
Db 480 LPRFAMDRAGLVGADGPTHGCAFDVTYMACLPNNVMAPSDEAEICMVAATAAIDRRPSCFRYPRG 539  
Qy 471 ASVRYPRGKGPAAIDPTL--TLEIGKAEVRHHSRIAILMGSVTAVRAGK----- 523

DB 540 SCFRPGRNGIGVQJL-PTGNKGTPLBEGKRLIEGERVALLGYSAGVONCLAAALSVLC 598  
QY 524 -GAGVAMNRFRKPPDQALVLELARTHVYFVVEENVVAGGASALNTPLOAKVL--- 579  
DB 599 HGLRLVADRFKPLDRSLIRSLANSHVLLIVEGSI--GGGSHVAQFMALDGLDCK 657  
QY 580 ---MPVCNIGLPDRFVQGSRELLSLVGLDSKGLIATI 615  
DB 658 LKMRPIV---LPDRYIDHGSPPADQSLAGITPSHIAIV 693

RESULT 6  
US-09-146-221-6

/ Sequence 6, Application US/09146221  
/ Patent No. 6190895  
/ GENERAL INFORMATION:  
/ APPLICANT: Croteau, Rodney B  
/ APPLICANT: Lange, Bernd M  
/ APPLICANT: Wildung, Mark R  
/ APPLICANT: McCaskill, David G  
/ TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el  
/ FILE REFERENCE: No. 6190895el, transketolase from pepper mint  
/ CURRENT APPLICATION NUMBER: US/09/146,221  
/ EARLIER FILING DATE: 1998-09-01  
/ EARLIER APPLICATION NUMBER: 60/056,033  
/ NUMBER OF SEQ ID NOS: 12  
/ SOFTWARE: Patentin Ver. 2.0  
/ SEQ ID NO 6  
/ LENGTH: 724  
/ TYPE: PRT  
/ ORGANISM: Mentha piperita  
US-09-146-221-6

Query Match 44.1%; Score 1418.5; DB 3; Length 724;  
Best Local Similarity 46.9%; Pred. No. 1.6e-133;  
Matches 298; Conservative 116; Mismatches 189; Indels 33; Gaps 13;

QY 8 PLKNIHTPADIRALSKDQLOLADVRGYLTHVTSISGHPAAGLITVLTALHYVEN 67  
DB 77 PILDITVYPMNKNLSVEELANLADLRBEIVTVSKTGHLSSLSVSELVLAHYVEN 136  
QY 68 TPVDQVMDVQVQAYPHKILTGKREMPITITLGVSAFPADSEYDAFVGHSSSTIS 127  
DB 137 TPDKRIIMVQHQAYPHKILTGKRAEMHTIRQTGLAGFPKDESAHDAFGAGHSSSTIS 196  
QY 128 AALGMAIASOLRGEDKQVVAIIGDSITGMAVEAMNHAGDVANLVLINDN--DMSI-- 184  
DB 197 AGLGMAVARDLQKNNHVISVIGDGMATGQAYEALNNGFLDSNLIIVANDNQVSLPT 256  
QY 185 -----SPVGMANNVLTIVLSKRYSSVRESKALAMPDS--VWEIARKTEBHVKMI- 236  
DB 257 ATVDGPAPVQALSKALITLQASRKFRQLREAAKSMTKOMGAPAHETASKLTQYVKGWNG 316  
QY 237 VPG--TLFEEIGFNYPCPIGDHVEMLVSTLENKDL--TGPFVLAHYVTKKGVAAPARKD 293  
DB 317 KEGASLFEELIGTYIGVDGNVEDLVYIFKAVKEMPAEGVLIHIIITKGGKGPAPARIA 376  
QY 294 PLAYHGVPAFPTDPLPKAAPSHPHYTVEFGRLCDMAADERLLGITPAMEGSGLV 353  
DB 378 ADMKGVVAFD--AKTGKQMKTKNKTSTYQYFASLVAEAEHDKIIVAHAMGGGTGLN 436  
QY 354 EFSQKFRNYFYDVAIAEOHVTTLAAGACGAPVVAIYSTFLORGVDOLIHVVALQNL 413  
DB 436 IFQKQPPDRCPDVGIAEOHVTTLAAGABGKLPCLAYSSFLORGVDVHVDVLDKLP 495  
QY 414 MLZALDRAGLVGPDPGTHAGAFDYSYMKCIPIMLIMAPADENECQML--TTGFOHHPAS 472  
DB 496 VRFVMDRAGVAGDGPTHGAFDPTTMACTPNNVVAAPSDAEALMNIATATAIIDRSPC 555  
QY 473 VRYRGGKGPAAIDPTL--TALFEGKAEVHHSRIILANGSMTPPAVBAKQL----- 525

DB 556 VRYRNGNGIGVALLPSNNKGTPLBEGKRLIEKSKVAILGSGTIVQNCMAANLLEQHI 615  
QY 526 GATVAMNFRKPPDQALVLELARTHVYFVVEENVVAGGASALNTPLOAKVL----- 579  
DB 616 SVTADAPFCRPLDGLIKLIVQEBRIVLIVEBSGI--GGFASHISFLINGLDGNLKW 674  
QY 580 MPVCNIGLPDRFVQGSRELLSLVGLDSKGLIATI 615  
DB 675 RPMV---LPDRYIDHGSPPADQSLAGITPSHIAIV 707

RESULT 7  
US-09-146-221-8

/ Sequence 8, Application US/09146221  
/ Patent No. 6190895  
/ GENERAL INFORMATION:  
/ APPLICANT: Croteau, Rodney B  
/ APPLICANT: Lange, Bernd M  
/ APPLICANT: Wildung, Mark R  
/ APPLICANT: McCaskill, David G  
/ TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el  
/ FILE REFERENCE: No. 6190895el, transketolase from pepper mint  
/ CURRENT APPLICATION NUMBER: US/09/146,221  
/ EARLIER FILING DATE: 1998-09-01  
/ EARLIER APPLICATION NUMBER: 60/056,033  
/ NUMBER OF SEQ ID NOS: 12  
/ SOFTWARE: Patentin Ver. 2.0  
/ SEQ ID NO 8  
/ LENGTH: 727  
/ TYPE: PRT  
/ ORGANISM: Mentha piperita  
US-09-146-221-8

Query Match 44.1%; Score 1418.5; DB 3; Length 727;  
Best Local Similarity 46.8%; Pred. No. 1.6e-133;  
Matches 298; Conservative 116; Mismatches 190; Indels 33; Gaps 13;

QY 8 PLKNIHTPADIRALSKDQLOLADVRGYLTHVTSISGHPAAGLITVLTALHYVEN 67  
DB 78 PILDITVYPMNKNLSVEELANLADLRBEIVTVSKTGHLSSLSVSELVLAHYVEN 137  
QY 68 TPVDQVMDVQVQAYPHKILTGKREMPITITLGVSAFPADSEYDAFVGHSSSTIS 127  
DB 138 TPDKRIIMVQHQAYPHKILTGKRAEMHTIRQTGLAGFPKDESAHDAFGAGHSSSTIS 197  
QY 128 AALGMAIASOLRGEDKQVVAIIGDSITGMAVEAMNHAGDVANLVLINDN--DMSI-- 184  
DB 198 AGLGMAVARDLQKNNHVISVIGDGMATGQAYEALNNGFLDSNLIIVANDNQVSLPT 257  
QY 185 -----SPVGMANNVLTIVLSKRYSSVRESKALAMPDS--VWEIARKTEBHVKMI- 236  
DB 258 ATVDGPAPVQALSKALITLQASRKFRQLREAAKSMTKOMGAPAHETASKLTQYVKGWNG 317  
QY 237 VPG--TLFEEIGFNYPCPIGDHVEMLVSTLENKDL--TGPFVLAHYVTKKGVAAPARKD 293  
DB 318 KEGASLFEELIGTYIGVDGNVEDLVYIFKAVKEMPAEGVLIHIIITKGGKGPAPARIA 377  
QY 294 PLAYHGVPAFPTDPLPKAAPSHPHYTVEFGRLCDMAADERLLGITPAMEGSGLV 353  
DB 378 ADMKGVVAFD--AKTGKQMKTKNKTSTYQYFASLVAEAEHDKIIVAHAMGGGTGLN 436  
QY 354 EFSQKFRNYFYDVAIAEOHVTTLAAGACGAPVVAIYSTFLORGVDOLIHVVALQNL 413  
DB 437 IFQKQPPDRCPDVGIAEOHVTTLAAGABGKLPCLAYSSFLORGVDVHVDVLDKLP 496  
QY 414 MLZALDRAGLVGPDPGTHAGAFDYSYMKCIPIMLIMAPADENECQML--TTGFOHHPAS 472  
DB 497 VRFVMDRAGVAGDGPTHGAFDPTTMACTPNNVVAAPSDAEALMNIATATAIIDRSPC 556  
QY 473 VRYRGGKGPAAIDPTL--TALFEGKAEVHHSRIILANGSMTPPAVBAKQL----- 525



Db 557 VRRPGRNGIGVALPNNKGTPLRIGKRIKESGKVALIGFTIYONCAAAATLEQHG1 616  
 Qy 526 GATVVMRRFPKPPDQALVLEARTHDVPTVEENVIA-GGAGSAINTELQAKVL----- 579  
 Db 617 STVADARCKPDLSDILKLVQEEVLTVEGSGIGIGSASHISHLISLNGLDN1K 676  
 Qy 580 -MPVCNIGLDRFVQSGREELSLVGLDSKGLIAT 615  
 Db 677 KRRPV---LPDRYIDHGAQSDQIEEAGLSPLHAGTV 710

RESULT 8

US-09-626-589-2  
 / Sequence 2, Application US/09626589  
 / Patent No. 6326164  
 / GENERAL INFORMATION:  
 / APPLICANT: Rice, John  
 / APPLICANT: Klotz, Andreas  
 / APPLICANT: Crawford, John  
 / APPLICANT: Lanning, Beth  
 / APPLICANT: Stewart, Sandy  
 / TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
 / IDENTIFICATION OF MODULATORS OF DEOXYXYLULOS 5-PHOSPHATE  
 / TITLE OF INVENTION: SYNTHASE ACTIVITY  
 / FILE REFERENCE: 2037 US  
 / CURRENT APPLICATION NUMBER: US/09/626,589  
 / CURRENT FILING DATE: 2000-07-27  
 / NUMBER OF SEQ ID NOS: 6  
 / SOFTWARE: FastSeq for Windows Version 4.0  
 / SEQ ID NO 2  
 / LENGTH: 659  
 / TYPE: PRT  
 / ORGANISM: Arabidopsis thaliana  
 / US-09-626-589-2

Query Match 43.9%; Score 1411.5; DB 4; Length 659;  
 Best Local Similarity 46.6%; Pred. No. 6.8e-133;  
 Matches 297; Conservative 115; Mismatches 188; Indels 37; Gaps 13;

Qy 8 PLKNIHTPADIRALSKDQLQDLADVRGYLTHTVVISGSGHFAAGLGTVELTVALHYFN 67  
 Db 17 PLDTITNYIHKMNLVSKLQSLDELRSDDVFNVSKTGHLGSSLGVELTVALHYFN 76  
 Qy 68 TPVDQLVMDVGHQAVPHKILTGKREKMPITRTLGVSAPPADESEYDAFCVGSSTIS 127  
 Db 77 TPQDKILMDVGHQSPHKLITGRKGMPTMRTGTNGISGPTKGESEHDCFTGHSSTIS 136  
 Qy 128 AALGMALASQLRGDKKVAIIIGDSITGMAVEAMNNAHDVNAVILINDN-DMSI-- 184  
 Db 137 AGLGMAVRGDLKGRNNVAVIIGDAMTAGQAYEAMNNAGYLDSDXVITLNDKQVSLPT 196  
 Qy 185 -----SPVGAANNVLTKLSSKRYSSVRESKALAMPSS-VWELARKTEEHVKMI- 236  
 Db 197 ATLDGSPFVGLASSLSRLQSNPALREIREVAKGKTQKQIGPMHQIAAKVDEVARGMIS 256  
 Qy 237 -VPGTLFRELGFNTFGSITGDHVEMLVSTLENLKD--LTGPVFLAVNTYKKGTAAPAKD 293  
 Db 257 GTGSSLFRELGLYYIGVDGNIDDLVALIKVKTGTGTPVLIHVYTERKRGYPVYERA 316  
 Qy 294 PLAYGVAPADPT--KDFLPKAPSPHPTYTEVFGWLCMAAODERLLGITPARRSGSG 351  
 Db 317 DDKXGVVKTDPARGRP---KTNKQOSTTYYFAELVVAEAVKDVVAIHAAMGGGTG 373  
 Qy 352 LVBSQKPPNRYFPAVAIAEONAVTLAAGACQAGAPVAIYSTLQNGYDOLIHVALON 411  
 Db 374 LNLFORPFPTRCFDVGIAEONAVTFAAGLACBGLKPCACAYSSFMORAYDQVVDVLDOK 433  
 Qy 412 LDMFALDRAGLVGDPGPTAGADVSYNRCIPNMLTMAPDEBEGCOMLTGFCQ-HHGP 470  
 Db 434 LPVAFADRAGLVGADGPTGCAFDVTFAACLPNMIWAPSDRALFKWATVAIADIRP 493  
 Qy 471 ASVRYPRGKPGAAIDP--TLTALTEIGKAEVRHHSRIAILAMGSMVTPAVEAG-----K 523

Db 494 SCFRYPGRNGIGVALPNNKGTPLRIGKRIKESGERVALLGYGSAVOSCLGAVALBER 553  
 Qy 524 QGATVVMRRFPKPPDQALVLEARTHDVPTVEENVIAAGAGSAINTELQAKVL----- 579  
 Db 554 GLMTVAARAFCKPDLRALIRSLAKSHEVLTVEGSGI-GGGSHVVOFLADGLDCKL 612  
 Qy 580 -MPVCNIGLDRFVQSGREELSLVGLDSKGLIAT 614  
 Db 613 KRRPV---LPDRYIDHGAQSDQIEEAGLSPLHAGTV 646

RESULT 9

US-09-626-589-1  
 / Sequence 1, Application US/09626589  
 / Patent No. 6326164  
 / GENERAL INFORMATION:  
 / APPLICANT: Rice, John  
 / APPLICANT: Klotz, Andreas  
 / APPLICANT: Crawford, John  
 / APPLICANT: Lanning, Beth  
 / APPLICANT: Stewart, Sandy  
 / TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
 / IDENTIFICATION OF MODULATORS OF DEOXYXYLULOS 5-PHOSPHATE  
 / TITLE OF INVENTION: SYNTHASE ACTIVITY  
 / FILE REFERENCE: 2037 US  
 / CURRENT APPLICATION NUMBER: US/09/626,589  
 / CURRENT FILING DATE: 2000-07-27  
 / NUMBER OF SEQ ID NOS: 6  
 / SOFTWARE: FastSeq for Windows Version 4.0  
 / SEQ ID NO 1  
 / LENGTH: 717  
 / TYPE: PRT  
 / ORGANISM: Arabidopsis thaliana  
 / US-09-626-589-1

Query Match 43.9%; Score 1411.5; DB 4; Length 717;  
 Best Local Similarity 46.6%; Pred. No. 7.8e-133;  
 Matches 297; Conservative 115; Mismatches 188; Indels 37; Gaps 13;

Qy 8 PLKNIHTPADIRALSKDQLQDLADVRGYLTHTVVISGSGHFAAGLGTVELTVALHYFN 67  
 Db 75 PLDTITNYIHKMNLVSKLQSLDELRSDDVFNVSKTGHLGSSLGVELTVALHYFN 134  
 Qy 68 TPVDQLVMDVGHQAVPHKILTGKREKMPITRTLGVSAPPADESEYDAFCVGSSTIS 127  
 Db 135 TPQDKILMDVGHQSPHKLITGRKGMPTMRTGTNGISGPTKGESEHDCFTGHSSTIS 194  
 Qy 128 AALGMALASQLRGDKKVAIIIGDSITGMAVEAMNNAHDVNAVILINDN-DMSI-- 184  
 Db 195 AGLGMAVRGDLKGRNNVAVIIGDAMTAGQAYEAMNNAGYLDSDXVITLNDKQVSLPT 254  
 Qy 185 -----SPVGAANNVLTKLSSKRYSSVRESKALAMPSS-VWELARKTEEHVKMI- 236  
 Db 255 ATLDGSPFVGLASSLSRLQSNPALREIREVAKGKTQKQIGPMHQIAAKVDEVARGMIS 314  
 Qy 237 -VPGTLFRELGFNTFGSITGDHVEMLVSTLENLKD--LTGPVFLAVNTYKKGTAAPAKD 293  
 Db 315 GTGSSLFRELGLYYIGVDGNIDDLVALIKVKTGTGTPVLIHVYTERKRGYPVYERA 374  
 Qy 294 PLAYGVAPADPT--KDFLPKAPSPHPTYTEVFGWLCMAAODERLLGITPARRSGSG 351  
 Db 375 DDKXGVVKTDPARGRP---KTNKQOSTTYYFAELVVAEAVKDVVAIHAAMGGGTG 431  
 Qy 352 LVBSQKPPNRYFPAVAIAEONAVTLAAGACQAGAPVAIYSTLQNGYDOLIHVALON 411  
 Db 432 LNLFORPFPTRCFDVGIAEONAVTFAAGLACBGLKPCACAYSSFMORAYDQVVDVLDOK 491  
 Qy 412 LDMFALDRAGLVGDPGPTAGADVSYNRCIPNMLTMAPDEBEGCOMLTGFCQ-HHGP 470  
 Db 492 LPVAFADRAGLVGADGPTGCAFDVTFAACLPNMIWAPSDRALFKWATVAIADIRP 551  
 Qy 471 ASVRYPRGKPGAAIDP--TLTALTEIGKAEVRHHSRIAILAMGSMVTPAVEAG-----K 523

Db 552 SCRRYPRNGIGVALPFGKGVPIEIGKRIKKEGRVALLGSAVSGCLGAAMLEBR 611  
 Qy 524 QLGATVNMRFVPEPDQALVLELARTHDVPEVVEENVAGAGSAINTFLOAKVY---- 579  
 Db 612 GLNVTADARFCCKPLRALIRLSAKSHEVLITVEBSI--GGFGSHVQVFLALDGLDGL 670  
 Qy 580 --MPVCNIGLPDRFVQSGREELSLVGLDSKGIAT 614  
 Db 671 KMRPMV--LPDRYIDHGAPADQLAEAGLMPSHIAT 704

RESULT 10  
 US-09-626-589-3

/ Sequence 3, Application US/09626589  
 / Patent No. 6326164  
 / GENERAL INFORMATION:  
 / APPLICANT: Rice, John  
 / APPLICANT: Klotz, Andreas  
 / APPLICANT: Crawford, John  
 / APPLICANT: Lanning, Beth  
 / APPLICANT: Stewart, Sandy  
 / TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
 / TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLOSE 5-PHOSPHATE  
 / FILE REFERENCE: 2037 US  
 / CURRENT APPLICATION NUMBER: US/09/626,589  
 / CURRENT FILING DATE: 2000-07-27  
 / NUMBER OF SEQ ID NOS: 6  
 / SOFTWARE: FASTSEQ for Windows Version 4.0  
 / SEQ ID NO 3  
 / LENGTH: 824  
 / TYPE: PRF  
 / ORGANISM: Artificial Sequence  
 / FEATURE:  
 / OTHER INFORMATION: Residues 1-165 are from the thioredoxin sequence  
 / OTHER INFORMATION: found in the vector pBT32 (supplied by No. 6326164agen.  
 / OTHER INFORMATION: Residues 166-824 represent the DXPS sequence from  
 / OTHER INFORMATION: Arabidopsis shown in SEQ ID NO:2.  
 US-09-626-589-3

Query Match 43.9%; Score 1411.5; DB 4; Length 824;  
 Best Local Similarity 46.6%; Pred. No. 9,9e-133;  
 Matches 297; Conservative 115; Mismatches 188; Indels 37; Gaps 13;

Qy 8 PLKNIHTPADIRALSKQLOQLADEVRGYLTHTVISISGHPAAGLCTVELTVALHYFN 67  
 Db 182 PLADTINYPHMKNLVVELKQLSDELRSDFVFNSTGHLGSSLGVELTVALHYFN 241  
 Qy 68 TPVDOLVWDVGHQAVPHKILITGRKERMPITRTIGVSAFPAPADESEYDAFGVGHSTIS 127  
 Db 242 TPQDKILMDVGHQSYFHKILITGRKMTPTMOTNGLSGFTKRGSEHDCFGTGHSTIS 301  
 Qy 128 AALGMAIASQLRGEDEKKNVAITIGDSTITGMAVEAMNHADVNANLITLNDN-DMSI- 184  
 Db 302 AGIGMAVAGDLKGNKNVAVITIGDAMTAGQAVEAMNNAAGYLDSDMITLINDNKQVSLPT 361  
 Qy 185 -----SPVGMANNVITKLSKFFYSVREESKKALAKMPS--VMEIARKEEYHVKMT- 236  
 Db 362 ATLDPSPVPGALSSALSRLOSNPLRZELREYAKGATKQGGPHEQLAAKVDETARKMTIS 421  
 Qy 237 -VPGTLFEELGFNYFGIDHDEMLVSTLENKD--LTGPVFLHVTYTKKGGYAPAEKD 293  
 Db 422 GTGSSLFEELGLVYIGVDSHNIDDLVAILKEVKSTRTGPLYLHVTEKRGYPYABRA 481  
 Qy 294 PLAYHGVAPADPT--KDFLPKAPSPHPTTYEVFGRMLCMAADDERLLGITPMRBSG 351  
 Db 482 DDKTHGVAKFDPATGRQF--KTTNKTSYTYFEBALVABAEVDKDVALLHAAAGGSGTG 538  
 Qy 352 LVEFSQKFPNRYFDVAIAEQHAVITLAAQACQAGAPVVAISTFLQSGYDOLIHDAVALON 411  
 Db 539 LNLFPQRFPTRCFDVGIAEQHAVITFAAGLACBGLPFCALYSSFMGRAYDQVVDVLDQX 598  
 Qy 412 LKMLPALDRAGLVGDPGPTAGAFDYSYMRCLPNMLIMAPADENECROMLTTGFCQ-HHGP 470

Db 599 LPVFPAMDRAGLVADGPTHGAFDVTFACTLPNMTVMAPEBDADEFMTATAVAIDRRP 658  
 Qy 471 ASVYPRCKGCAIDP--TITALEIGAPVRHHGSRITALLAKMSMTTPABAG-----K 523  
 Db 659 SCFRPFRNGIGVALPFGKGVPIEIGKRIKKEGRVALLGSAVSGCLGAAMLEBR 718  
 Qy 524 QLGATVNMRFVPEPDQALVLELARTHDVPEVVEENVAGAGSAINTFLOAKVY---- 579  
 Db 719 GLNVTADARFCCKPLRALIRLSAKSHEVLITVEBSI--GGFGSHVQVFLALDGLDGL 777  
 Qy 580 --MPVCNIGLPDRFVQSGREELSLVGLDSKGIAT 614  
 Db 778 KMRPMV--LPDRYIDHGAPADQLAEAGLMPSHIAT 811

RESULT 11

US-09-857-556A-10  
 / Sequence 10, Application US/09857556A  
 / Patent No. 6558915  
 / GENERAL INFORMATION:  
 / APPLICANT: Rebecca E. Cahoon  
 / APPLICANT: Sean J. Coughlan  
 / APPLICANT: Yong Tao  
 / APPLICANT: Zude Keng  
 / APPLICANT: Mark E. Williams  
 / TITLE OF INVENTION: Plant 1-Deoxy-Xylose 5-Phosphate Synthase  
 / FILE REFERENCE: BB1290  
 / CURRENT APPLICATION NUMBER: US/09/857,556A  
 / CURRENT FILING DATE: 2001-06-04  
 / PRIOR APPLICATION NUMBER: 60/1110,779  
 / PRIOR FILING DATE: 1998-12-03  
 / NUMBER OF SEQ ID NOS: 34  
 / SOFTWARE: Microsoft Office 97  
 / SEQ ID NO 10  
 / LENGTH: 721  
 / TYPE: PRF  
 / ORGANISM: Glycine max  
 US-09-857-556A-10

Query Match 43.9%; Score 1410.5; DB 4; Length 721;  
 Best Local Similarity 47.0%; Pred. No. 1e-132;  
 Matches 302; Conservative 107; Mismatches 188; Indels 45; Gaps 14;

Qy 8 PLKNIHTPADIRALSKQLOQLADEVRGYLTHTVISISGHPAAGLCTVELTVALHYFN 67  
 Db 76 PLADTINYPHMKNLVVELKQLSDELRSDFVFNSTGHLGSSLGVELTVALHYFN 135  
 Qy 68 TPVDOLVWDVGHQAVPHKILITGRKERMPITRTIGVSAFPAPADESEYDAFGVGHSTIS 127  
 Db 136 APQDKILMDVGHQSYFHKILITGRKMTPTMOTNGLSGFTKRGSEHDCFGTGHSTIS 195  
 Qy 128 AALGMAIASQLRGEDEKKNVAITIGDSTITGMAVEAMNHADVNANLITLNDN-DMSI- 185  
 Db 196 AGIGMAVAGDLKGNKNVAVITIGDAMTAGQAVEAMNNAAGYLDSDMITLINDNKQVSLPT 255  
 Qy 186 -----SPVGMANNVITKLSKFFYSVREESKKALAKMPS--VMEIARKEEYHVKMT- 236  
 Db 256 ANLDGPVPGALSSALSRLOSNPLRZELREYAKGATKQGGPHEQLAAKVDETARKMTIS 315  
 Qy 237 -VPGTLFEELGFNYFGIDHDEMLVSTLENKD--LTGPVFLHVTYTKKGGYAPAEKD 293  
 Db 316 GSGSTLFEELGLVYIGVDSHNIDDLVAILKEVKSTRTGPLYLHVTEKRGYPYABRA 375  
 Qy 294 PLAYHGVAPADPT--KDFLPKAPSPHPTTYEVFGRMLCMAADDERLLGITPMRBSG 351  
 Db 376 ADKTHGVAKFDPATGRQF--KTTNKTSYTYFEBALVABAEVDKDVALLHAAAGGSGTG 432  
 Qy 352 LVEFSQKFPNRYFDVAIAEQHAVITLAAQACQAGAPVVAISTFLQSGYDOLIHDAVALON 411  
 Db 433 LNLFPQRFPTRCFDVGIAEQHAVITFAAGLACBGLPFCALYSSFMGRAYDQVVDVLDQX 492  
 Qy 412 LKMLPALDRAGLVGDPGPTAGAFDYSYMRCLPNMLIMAPADENECROMLTTGFCQ-HHGP 470

```

Db 493 LPVRAMRAGLVCAGDGPTHGCAFDVTNACLPNNVYVMAPODEBELFHVATAAIDRRP 552
Qy 471 ASVRYRGRGKRAAID-----PTLALSIGKAVNHHSSRLAIIAAGSMTPAVEAGK- 523
Db 553 SCFRYPGRNGIGVELPAGKNGIP---LEIGKGRILIEBERVALLGYSAGVONCLAAASL 608
Qy 524 ----OLGATVVMREYKPPDQALVLELARTHDVFTVBERNVLAGAGSAINTELOAKYL 579
Db 609 LHHGGRATVADARFCXKEDRSILRSLSAOSHEVLIVBERGSI--GGFGSHVQFMALDGL 667
Qy 580 -----MPVCNIGLPDRFVEQSGREELSLVGLDSKGIAT 615
Db 668 DKLKMRPIV---LPDCYIDHSGSPVDQLSAGLTPSHIATV 706

```

## RESULT 12

```

US-09-857-556A-33
; Sequence 33, Application US/09857556A
; Patent No. 6558915
; GENERAL INFORMATION:
; APPLICANT: Rebecca B. Cahoon
; APPLICANT: Sean J. Coughlan
; APPLICANT: Yong Tao
; APPLICANT: Zude Weng
; APPLICANT: Mark E. Williams
; TITLE OF INVENTION: Plant 1-Deoxy-Xyloiose 5-Phosphate Synthase
; FILE REFERENCE: BH1290
; CURRENT APPLICATION NUMBER: US/09/857,556A
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/110,779
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
; LENGTH: 719
; TYPE: prt
; ORGANISM: Capsicum annuum
US-09-857-556A-33

```

```

Query Match 43.5%; Score 1397.5; DB 4; Length 719;
Best Local Similarity 46.1%; Pred. No. 2e-111;
Matches 294; Conservative 114; Mismatches 193; Indels 37; Gaps 13;

```

```

Qy 8 PLEKNIHTPADIRALSKDQLQALADEVRGYLTHTVSISGGHFAAGLGTVELTVALHYFN 67
Db 74 PIVDTINYPIMHKNSLKEKQLADELRSDITFNVSKTGHLGSLGVVETLVALHYFN 133
Qy 68 TPVQQLVWDVGHQAYPHKILITGRKEMPTITRLGVSAPFAPADESEYDAFGVGHSSSTIS 127
Db 134 APQDRLTMDVGHQSTPHKILITGRKEMPTITRLGVSAPFAPADESEYDAFGVGHSSSTIS 193
Qy 128 AALGMAIASOLRGEDKKVVAIIIGDSITGGAAVEAMNAGDVANLILVINDN--DMSIS- 185
Db 194 AGLGMAVARDLKGKNNNYIAVIGDGMATAGQAYEAMNAGVILDSMTIILINDNQVSLPT 253
Qy 186 -----PPVGAMNNYITKVLSKFFYSVRESKALAKMPS--WELARKTEEHYKMI- 236
Db 254 ATLDPGPVPVAGLSALSRLQSNRPLREIRAKVTKQIGGPMHEILAAXKDEVARQWIS 313
Qy 237 -VPGTLFPELGNVYGPIDGHDVEMLVSTLEKLD--LTGVPFLHVYTKKGGKGAAPAKD 293
Db 314 GSGSTLFEELGITYIGPVDGNIDILISILEKVSSTKTTGVPVLHVTEKRGKGPYIAERA 373
Qy 294 PLAYHGVAPDPT--KDFLPKAAAPSPHTYTEVEFGMWLCDMAADERLLGITPPARESGG 351
Db 374 ADKHGVAKFDPATGKQPKSGAKTOSYTTY--FAEALIAABADKDIVALIHAAAGCGTG 430
Qy 352 LVSQKPRNRYEPDAIEQHAUTLAAGQACGAPVVAISTLQCGYDQLIHVDALON 411
Db 431 KMLFLKRRPTFCFDGIAEQHAUTLAAGLACGELPFCALISSPQGRAYDDVADVLOK 490
Qy 412 LDMFLALDRAGLVGPDGPTHAGAFDYSTYRCIPNNLIMAPADENECROMLTTGQ--HHGP 470

```

```

Db 491 LPVRAMRAGLVCAGDGPTHGCAFDVTNACLPNNVYVMAPODEBELFHVATAAIDRRP 550
Qy 471 ASVRYRGRGKRAAIDPTLTA--LEIGRAVEHHSRLAIIAAGSMTPAVEA-----GK 523
Db 551 SCFRYPGRNGIGVELPAGKNGIP---LEIGKGRILIEBERVALLGYSAGVONCLAAASL 610
Qy 524 ----OLGATVVMREYKPPDQALVLELARTHDVFTVBERNVLAGAGSAINTELOAKYL----- 579
Db 611 GLQYTVADARFCXKEDRSILRSLSAOSHEVLIVBERGSI--GGFGSHVQFMALDGLDKL 669
Qy 580 -----MPVCNIGLPDRFVEQSGREELSLVGLDSKGIAT 615
Db 670 KMRPIV---LPDRYIDHSGSPVDQLSAGLTPSHIATV 704

```

## RESULT 13

```

US-09-146-221-4
; Sequence 4, Application US/09146221
; Patent No. 6190895
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; APPLICANT: Wildung, Mark R
; APPLICANT: McCaskill, David G
; TITLE OF INVENTION: Nucleic and Amino Acid Sequences Relating to a No. 6190895el
; FILE REFERENCE: No. 6190895el transketolase from peppermint
; CURRENT FILING DATE: 1998-09-01
; EARLIER APPLICATION NUMBER: 60/056,033
; PRIOR FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 721
; TYPE: prt
; ORGANISM: Mentha piperita
US-09-146-221-4

```

```

Query Match 43.3%; Score 1392; DB 3; Length 721;
Best Local Similarity 46.5%; Pred. No. 7.2e-121;
Matches 296; Conservative 115; Mismatches 189; Indels 36; Gaps 14;

```

```

Qy 8 PLEKNIHTPADIRALSKDQLQALADEVRGYLTHTVSISGGHFAAGLGTVELTVALHYFN 67
Db 77 PIVDTINYPIMHKNSLKEKQLADELRSDITFNVSKTGHLGSLGVVETLVALHYFN 136
Qy 68 TPVQQLVWDVGHQAYPHKILITGRKEMPTITRLGVSAPFAPADESEYDAFGVGHSSSTIS 127
Db 137 TPQDKITMDVGHQAYPHKILITGRKEMPTITRLGVSAPFAPADESEYDAFGVGHSSSTIS 196
Qy 128 AALGMAIASOLRGEDKKVVAIIIGDSITGGAAVEAMNAGDVANLILVINDN--DMSI- 184
Db 197 AGLGMAVARDLKGKNNNYIAVIGDGMATAGQAYEAMNAGVILDSMTIILINDNQVSLPT 256
Qy 185 -----SPVGAMNNYITKVLSKFFYSVRESKALAKMPS--WELARKTEEHYKMI- 236
Db 257 ATVDGPAPVAGLSALTKLQASRKFQRLREAAKSMTKGAGAHBAHSLTLTVYKGMWG 316
Qy 237 VPG--TLPEHGFVYFGIDHDVEMLVSTLEKLD--TPVPVFLHVYTKKGGKGAAPAKD 293
Db 317 KPGASLEBEHIGITYIGV--DVEDLVYIFKVKEMPAEPVLIHITTEKRGKGPYIAERA 373
Qy 294 PLAYHGVAPDPTKDFLPKAAAPSPHTYTEVEFGMWLCDMAADERLLGITPPARESGGLV 351
Db 374 ADKHGVAKFDPATGKQPKSGAKTOSYTTY--FAEALIAABADKDIVALIHAAAGCGTG 432
Qy 354 EESQKPRNRYEPDAIEQHAUTLAAGQACGAPVVAISTLQCGYDQLIHVDALON 413
Db 433 IFQKQPRDRCTFDGIAEQHAUTLAAGLACGELPFCALISSPQGRAYDDVADVLOK 492
Qy 414 MLFLALDRAGLVGPDGPTHAGAFDYSTYRCIPNNLIMAPADENECROMLTTGQ--HHGP 472

```

```

Db      493 VAPMDRAGVADGTHGADPTTMACTLPNNVWAPDEBELMMATAIIDR89C 552
Qy      473 VAYPRKGPAAIDPTL--TALEIGAEVREHSGRIAILAMGSWTPAVEAGKQI----- 525
Db      553 VAYPRKNGIGVALPSNNKGTPLEIGKRIKESKVALIGFTIYQNCMAANLLEHQGI 612
Qy      526 GATVVMREVRKPPDQALVIELARTHDVFTVEENVYAGAGSAINTELQAKVU----- 579
Db      613 SYTVADARCKPDLGDLIKKLVOHESVLITVEGSI--GGPSAHISHLISLNGLDGNLKW 671
Qy      580 MPVCNIGLPDRFEVCGREELSLVGLDSKGIAT 615
Db      672 RMMV---LPDRYIDHGAQSDQIEBAGLSKHIAGTV 704

RESULT 14
US-09-198-452A-1135
; Sequence 1135, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1135
; LENGTH: 644
; TYPE: PR
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-1135

Query Match      35.7%; Score 1148; DB 4; Length 644;
Best Local Similarity 40.2%; Pred. No. 2.1e-106;
Matches 249; Conservative 115; Mismatches 236; Indels 20; Gaps 8;

Qy      8 PLTKNHTPADIRALSKDQLOQLADVRGYLTHTYISGGHFAAGLGYTELVTALHYEN 67
Db      7 PLIDILSRADLKKSISQIPGLAEIRIRIISVLSOTGHLSSNLGIVELTALHYPS 66
Qy      68 TPVDQIWMVGHQAYPHKILTGR--KERMTIRTLGGVSAFPARDESEYDAFGVSHSSTSI 126
Db      67 SPDKXIFIVGHQTYPHKILTGRNNGFPHINDNGLSGFTNPTSDHDLFPGAGATL 126
Qy      127 SAALGMAISQIRGEBKQVAILIGDSITGGMAYEAMHAKGVNANLVTILANDMSISP 186
Db      127 SIALGMAQTTPLESR--THVIFILGDAFSCGLTEALNNISITDLSKFVYILANDNMSISK 185
Qy      187 PVGAMNNYTLTKVLSKFYSVREESKKALAKP---SWELARKTEEHVKMTIVGTLF 242
Db      186 NYGAMRISIRSMHHTATKLTQVYEKMAKIPRYDSILAKSRSLQCVKILFCTPLF 245
Qy      243 EELGENYFPGPIDGHVEMLVSTLENKDLTGPVFLHVTYKKGATAPAEKDPALAHGVPA 302
Db      246 EPGGLAYVGPIDSHNYKLIPILOSVYRNLPPFVLHVCTTKKGLDQANNPAKHYGVA 305
Qy      303 -FPPTDPLPKAASPHTYTEVFGKMLDMMAAQDBRLIGTIPABEGSGLVAFQKXPN 361
Db      306 NPKKESAGHLPIKXPSFPDIFGQTLCELGVSRLHVVTPMSSIGSLKGFQKXPE 365
Qy      362 RYFDVALAEQAVTLAAGACOGAKFVAIYSTFLRGYDOLIHVALONTMLPALDRA 421
Db      366 RFDVGIAGHAYTFSAGIAKAGNPVICSISTFLRALDNYFDVCOMDLFVIFALDRA 425
Qy      422 GLVGPQCPHAGAFDYSVWCIPKMLIMAPADENECQMLTGFQHGASASRYRKGKP 481
Db      426 GLAYGGRSHHGILYDMSFLAMPOKITCOPRSQVVFQOULLYSSLHSSSALRIPIAP 485
Qy      482 GAALDPTLTA-----LEIGAEVREHSGRIAILAMGSWTPAVEAGKQI-----GATVN 531

```

```

Db      486 HG--DP-LTGDPNRLRSPGNAFTLSQGBDVILIALGLTCTALSIKQOLAYGISATVVD 542
Qy      532 MRVYPPDQALVELFARTHDVFTVEENVYAGAGSALNTELQAKVIMPCNIGLPDR 591
Db      543 PIFKPPDNDLFFSLIMSHSKVITIEBSHIRGIASSEFNVAATFNKVDILNPAIDTF 602
Qy      592 VEOGSEELSLVGLDSKI 611
Db      603 LSHGSKHALTYSIGLDESSK 622

RESULT 15
US-08-311-731A-56
; Sequence 56, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESS: WOLF, GREENFIELD & SACKS, P. C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 736 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIUM LAPRAE
US-08-311-731A-56

Query Match      33.8%; Score 1086.5; DB 4; Length 736;
Best Local Similarity 38.1%; Pred. No. 4.1e-100;
Matches 237; Conservative 119; Mismatches 245; Indels 21; Gaps 9;

Qy      9 LKKNHTPADIRALSKDQLOQLADVRGYLTHTYISGGHFAAGLGYTELVTALHYEN 68
Db      94 MLBOIRRRADLQHSQOQLRDLALIEBLVHKVAAATGGLGPNVGLVETLALHRYFDS 153
Qy      69 PVDDQIWMVGHQAYPHKILTGRKERMTIRTLGGVSAFPARDESEYDAFGVSHSSTSI 128
Db      154 PHDPIPTGTHQAYVHKILTGRQDPQSLRKKAAGLSGVPRAESEHVDVBSHSTALSY 213
Qy      129 ALGMAISQIRGEBKQVAILIGDSITGGMAYEAMHAKGVNANLVTILANDMSISP 187
Db      214 ADGLAKAFELAGNNRRHVAVVAGDALTGCMCEALNNIATPRPVYIVANDNGRSYAF 273
Qy      188 VGAMNNYTLTKVLSKFYSVREESKKALAKP--SWELARKTEEHVKMTIVGTLF 243

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 18.2424 Seconds  
(without alignments) 3268.453 Million cell updates/sec

Title: US-09-941-947a-6

Perfect score: 3216

Sequence: 1 MKLTIDYELAKNIHHPADIR.....LSVGLSDSKGLIATIEQPCA 620

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARYS

| Result No. | Score  | Query Match | Length DB | ID       | Description            |
|------------|--------|-------------|-----------|----------|------------------------|
| 1          | 2097.5 | 65.2        | 626       | 2 H82266 | 1-deoxyxylinulose-5-   |
| 2          | 2069.5 | 64.4        | 627       | 2 G83139 | 1-deoxyxylinulose-5-   |
| 3          | 2015   | 62.7        | 619       | 2 A10385 | 1-deoxy-D-xylinulose   |
| 4          | 2014.5 | 62.6        | 620       | 2 AF0554 | 1-deoxyxylinulose-5-   |
| 5          | 2000.5 | 62.2        | 620       | 2 D64771 | dxs protein - Esch     |
| 6          | 1995.5 | 62.0        | 620       | 2 B90688 | 1-deoxy-D-xylinulose   |
| 7          | 1995.5 | 62.0        | 620       | 2 F85538 | 1-deoxy-D-xylinulose   |
| 8          | 1951   | 60.7        | 620       | 2 A82582 | 1-deoxyxylinulose-5-ph |
| 9          | 1937.5 | 60.2        | 625       | 2 B64172 | dxs protein - Haem     |
| 10         | 1776   | 55.2        | 637       | 2 B81978 | probable 1-deoxyxy     |
| 11         | 1775   | 55.2        | 637       | 2 D81034 | 1-deoxyxylinulose-5-   |
| 12         | 1667   | 51.8        | 608       | 2 A84984 | dxs protein (import    |
| 13         | 1620.5 | 49.5        | 628       | 2 A70376 | conserved hypothet     |
| 14         | 1591.5 | 49.5        | 643       | 2 AD3339 | 1-deoxyxylinulose-5-   |
| 15         | 1523.5 | 47.4        | 640       | 2 C87505 | 1-deoxyxylinulose-5-   |
| 16         | 1499.5 | 46.6        | 639       | 2 A97450 | 1-deoxy-D-xylinulose   |
| 17         | 1499.5 | 46.6        | 639       | 2 AC2668 | 1-deoxy-D-xylinulose   |
| 18         | 1459   | 45.4        | 629       | 2 C83997 | 1-deoxyxylinulose-5-   |
| 19         | 1452   | 44.8        | 629       | 2 S75175 | hypothetical prote     |
| 20         | 1434   | 44.6        | 635       | 2 AF1881 | 1-deoxy-xylinulose 5   |
| 21         | 1428   | 44.4        | 634       | 2 T02208 | transketolase-like     |
| 22         | 1411.5 | 43.9        | 717       | 2 H85171 | DBF (CNA1) protein     |
| 23         | 1410.5 | 43.9        | 641       | 2 G28771 | hypothetical prote     |
| 24         | 1402.5 | 43.6        | 739       | 2 T52289 | probable transke       |
| 25         | 1402   | 43.6        | 735       | 2 T08140 | 1-deoxy-D-xylinulose   |
| 26         | 1400   | 43.5        | 632       | 2 A11607 | D-1-deoxyxylinulose    |
| 27         | 1397.5 | 43.5        | 719       | 2 T09543 | deoxyxylinulose synt   |
| 28         | 1394.5 | 43.4        | 633       | 2 B69961 | conserved hypothet     |
| 29         | 1366.5 | 42.5        | 619       | 2 A97156 | deoxyxylinulose-5-ph   |

|    |        |      |     |          |                      |
|----|--------|------|-----|----------|----------------------|
| 30 | 1351   | 42.0 | 609 | 2 AE1245 | D-1-deoxyxylinulose  |
| 31 | 1326   | 41.2 | 629 | 2 G75390 | 1-deoxy-D-xylinulose |
| 32 | 1274.5 | 39.6 | 608 | 2 A72213 | 1-deoxyxylinulose-5- |
| 33 | 1259.5 | 39.2 | 703 | 2 D71420 | hypothetical prote   |
| 34 | 1227.5 | 38.2 | 630 | 2 B71276 | probable transke     |
| 35 | 1227   | 38.2 | 615 | 2 E81451 | 1-deoxyxylinulose-5- |
| 36 | 1148   | 35.7 | 644 | 2 A86623 | transketolase (imp   |
| 37 | 1148   | 35.7 | 644 | 2 A72002 | 1-deoxyxylinulose-5- |
| 38 | 1119.5 | 34.8 | 656 | 2 T35408 | probable transke     |
| 39 | 1106.5 | 34.4 | 618 | 2 H71946 | 1-deoxyxylinulose-5- |
| 40 | 1099.5 | 34.2 | 618 | 2 B64564 | transketolase B -    |
| 41 | 1087.5 | 33.8 | 638 | 2 E70528 | probable dxs prote   |
| 42 | 1086.5 | 33.8 | 643 | 2 H87038 | hypothetical prote   |
| 43 | 1072.5 | 33.3 | 632 | 2 E81684 | 1-deoxyxylinulose-5- |
| 44 | 1044.5 | 32.5 | 640 | 2 F71527 | probable transke     |
| 45 | 886    | 27.5 | 580 | 2 A68354 | 1-deoxyxylinulose-5- |

#### ALIGNMENTS

##### RESULT 1

H82266  
1-deoxyxylinulose-5-phosphate synthase VC0889 [imported] - Vibrio cholerae (strain N16961)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: H82266  
R:Heidelberg, J.F.; Bisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
cardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Base, S.; Qin, H.; Dragot, I.; Sellers, P.  
1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.W.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: H82266  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-626 <H82>  
A:Cross-references: GB:AE004173; GB:AE003852; NID:59655341; PIDN:AAF94051.1; GSPDB:GN001;  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
A:Genetics:  
A:Gene: VC0889  
A:Map position: 1  
C:Superfamily: hypothetical protein C2814

|                       |       |  |                 |                   |
|-----------------------|-------|--|-----------------|-------------------|
| Query Match           | 65.2% | Score 2097.5;  | DB 2;           | Length 626;       |
| Best Local Similarity | 62.5% | Pred. No. 2.8e-139;  |                 |                   |
| Matches               | 391;  | Conservative 109;  | Mismatches 117; | Indels 9; Gaps 3; |
| QY                    | 3     | LTITD---YPLAKNIHTPADIRALSKDQLQCLADVRGILTTTSTSGCHFAAGTVELT  | 59              |                   |
| DB                    | 1     | MTLIDSKYPTLALANTPDELRLPREVLPKLCDELRITLINSVSQSSCHLASGLTVELT | 60              |                   |
| QY                    | 60    | VALATVFTVPDQVWDVGHQAYPHKILTGKREMERITRTIGVSAPFAPRDESBYDARGV | 119             |                   |
| DB                    | 61    | VALATVFTVPDQVWDVGHQAYPHKILTGKREMERITRTIGVSAPFAPRDESBYDARGV | 120             |                   |
| QY                    | 120   | GHSTSTISAAIGMAIASQLRGEDKKMAIIDDSTIGMAIYEAHNAHADVNAALLVIND  | 179             |                   |
| DB                    | 121   | GHSTSTISAAIGMAIACAKGEGDKRVSVIDGATTAAGAFRANNAHADVNAALLVIND  | 180             |                   |
| QY                    | 180   | NDMSTISPVGAMNNVLTIVLSKFFSVREESKALALAPSVBELARTSEHVKMIVPG    | 239             |                   |
| DB                    | 181   | NEMSTISPVGALNNHNLAVLSGLSYTSIREGGKKVSGIPIKELVYRTSEHLDGMVVPG | 240             |                   |
| QY                    | 240   | TLFEELGNNYFOPIDGHVEMVSTLENLKDVLTPGVFLHVVTKKGGYAPAKEDPLAHG  | 299             |                   |
| DB                    | 241   | TLFEELGNNYFOPIDGHVLELITLKNRRELKGPQFLHVTATKGGIYAPAKEDPIGHHG | 300             |                   |
| QY                    | 300   | VPAFPTKDFLPKAPSPHPTTETVFGRLCDMAAODERLGITPAPMEGSGLVESQKF    | 359             |                   |
| DB                    | 301   | VKFPDPSHSLPKSS-NTKQPTFSKIFDPLCDMAAODPKLAITPAPMEGSGMVRFSKEY | 359             |                   |
| QY                    | 360   | PNRYEDVAIAEOHVAITLAAGAACGAKPVAIYSTFELQGVQDLIHVAQLNDMLPALD  | 419             |                   |

```

Db      360 PQQYEVVAIAEQAHTVLAAGMAIAGHPVIAVSTFLOQGYDOLIHDAVIMPLPMPFAID 419
Qy      420 RAGLVDPDGTAGADYVYMCIPNMLIMAPADENECOMLTGTFQHHGPASVAYPPGX 479
Db      420 RAGIVADQOTHOAGFADLSYMCIPNMLIMAPADENECOMLTGTFQHHGPASVAYPPGN 479
Qy      480 GFGAIDPTLTAIEIGKAEVRH-----HGSRJATIAMGSMVTPAYAGKOLATVNNRF 534
Db      480 GMSVELESFTALIEIGKRLMRSTACEGEKVAIISFTLLPNAIOAKRNATVADNR 539
Qy      535 VMPEDQALVLEARTDVFVTEENVIAGGASALNTFQAOKVIMPCNIGLPRFPEBQ 594
Db      540 VEPLEBALIKOLAQTHDVLTLEENAIAGGAGVIEFLMKERKOLKPVNLGLPDPQTVQ 599
Qy      595 GSRBELLSVGLDSKGIATIBQFCA 620
Db      600 GTOEWHALGLDGAIGIERAIDVLA 625

```

## RESULT 2

```

G83139
1-deoxy-5-phosphate synthase PA4044 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83139
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lam,
.; Lory, S.; Olson, M.V.
Mature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:2043737; PMID:10984043
A:Accession: G83139
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <STO>
A:Cross-References: GB:AE004821; GB:AE004091; NID:99950236; PIDN:AA07431.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: dks; PA4044
C:Superfamily: hypothetical protein C2814

```

```

Query Match      64.4%; Score 2069.5; DB 2; Length 627;
Best Local Similarity 63.4%; Pred. No. 2.6e-137;
Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;

```

```

Qy      8 PILLKNIHPADIRALSKQLOQLADEVRGYLTHVTSISGHPFAAGLGVETLVLAHYVN 67
Db      16 PLMDRASSFAELRRLEADLETLADELKOYLLVYGQTHGFGAGLGVETLIALHYVD 75
Qy      68 TEPDQLVMDVGHQAVPHKILTGKERMPTIRTLGVSAFPARDSEYDAFGVGHSTIS 127
Db      76 TEPDRLVMDVGHQAVPHKILTERRELMTLRQKNGLAAPPRASEYDTPFGVGHSTIS 135
Qy      128 PALLGAIASQLGDEKQKVAIIGDSITGMA YEAMNHAQDVANLVTINDNMSISPP 187
Db      136 PALLGAIASQLGDEKQKVAIIGDSITGMA FEALNHAQDVADMLVTINDNMSISHN 195
Qy      188 VGMANNVLTKVLSKFSYSSVREESKKALAKMPSVVELARKTEEHVKGMIVPGTLFEELG 247
Db      196 VGLLSNLTAKILSSRTYSMRGSKKVALSLPGAMEIARRTEEVAKMGLVPGTLFEELG 255
Qy      248 NYFGPIDGHVEMVSTLENKDLTGPFVFLHVTYKKGKYPAPAEKDPALHGVPAEDPTX 307
Db      256 NYIGPIDGHVEMVSTLENKDLTGPFVFLHVTYKKGKYPAPAEKDPALHGVPAEDPTX 310
Qy      308 DFLPAPAP--SPHPTVTEVFGKMLCDMAADERLLGITPPAMREGSGLVESQKPNRYFD 365
Db      311 LEAPSSAPKKTGGPRYSVTFGQMLCDMAADARLLGITPPAMREGSDVAFSERYERKFD 370
Qy      366 VAIAGQHAVTTAAAGQACGAKRYVAIISTFLORGYDOLIHVAIONLMDLPAIRAGLVG 425
Db      371 VAIAGQHAVTTAAAGQACGAKRYVAIISTFLORGYDOLIHVAIONLMDLPAIRAGLVG 430

```

```

Qy      426 PDGPTHAGAPYSYMCIPNMLIMAPADENECOMLTGTFQHHGPASVAYPPGCAAI 485
Db      431 EDGPTHAGAPYSYMCIPNMLIMAPADENECOMLTGTFQHHGPASVAYPPGCAAI 490
Qy      486 DPTLTALIEIGKAEVRH-----HGSRJATIAMGSMVTPAYAGKOLATVNNRF 545
Db      491 DPTLTALIEIGKAEVRH-----HGSRJATIAMGSMVTPAYAGKOLATVNNRF 550
Qy      546 LARTDVFVTEENVIAGGASALNTFQAOKVIMPCNIGLPRFPEBQ 605
Db      551 LAGSHELLVTEENVIAGGASALNTFQAOKVIMPCNIGLPRFPEBQ 610
Qy      606 LDSKGIATIBQ 617
Db      611 LDAAGIERAVRQ 622

```

## RESULT 3

```

A10385
1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.-.-) [imported] - Yersinia pestis (strai
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: A10385
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibbally, R.W.; Hoiden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
H., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Mature 413, 523-527, 2001
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: A10385
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-619 <KIR>
A:Cross-References: GB:AF590842; PIDN:CA92412.1; PID:91598115; GSPDB:GN00175
A:Genetics:
A:Gene: dks
C:Superfamily: hypothetical protein C2814
C:Keywords: transferase

```

```

Query Match      62.7%; Score 2015; DB 2; Length 619;
Best Local Similarity 61.6%; Pred. No. 1.7e-133;
Matches 378; Conservative 95; Mismatches 139; Indels 2; Gaps 2;

```

```

Qy      7 YPLKNIHPADIRALSKQLOQLADEVRGYLTHVTSISGHPFAAGLGVETLVLAHYVN 66
Db      8 YPLTALAEHBEELKMLPKESLPKLCBRLKOYLLTVCSRSSGHRASGLGVETLIALHYV 67
Qy      67 NTPVDQLVMDVGHQAVPHKILTGKERMPTIRTLGVSAFPARDSEYDAFGVGHSTIS 126
Db      68 NTPVDHLMVMDVGHQAVPHKILTGKERMPTIRTLGVSAFPARDSEYDAFGVGHSTIS 127
Qy      127 SAALGMAIASQLGDEKQKVAIIGDSITGMA YEAMNHAQDVANLVTINDNMSISPP 186
Db      128 SAALGMAIASQLGDEKQKVAIIGDSITGMA FEAMNHAQDVADMLVTINDNMSIS 187
Qy      187 PYGMANNVLTKVLSKFSYSSVREESKKALAKMPSVVELARKTEEHVKGMIVPGTLFEELG 246
Db      188 NYGGINNHLKQLLSGKLYALREGGKAPALPPIKOLKRTTEHLLKGMVVPSTLFEELG 247
Qy      247 FNYFGPIDGHVEMVSTLENKDLTGPFVFLHVTYKKGKYPAPAEKDPALHGVPAEDPT 306
Db      248 FNYFGPIDGHVEMVSTLENKDLTGPFVFLHVTYKKGKYPAPAEKDPALHGVPAEDPT 307
Qy      307 KDPLPKAPSPHPTVTEVFGKMLCDMAADERLLGITPPAMREGSGLVESQKPNRYFDV 366
Db      308 SGLTPK--SQSLPTLYSKIFGEMLCETPAKOSKIMAVTPAMREGSGMVFREYPOQYFDV 366
Qy      367 AIAEGHAVTTAAAGQACGAKRYVAIISTFLORGYDOLIHVAIONLMDLPAIRAGLVG 426
Db      367 AIAEGHAVTTAAAGQACGAKRYVAIISTFLORGYDOLIHVAIONLMDLPAIRAGLVG 426
Qy      427 DGPTHAGAPYSYMCIPNMLIMAPADENECOMLTGTFQHHGPASVAYPPGCAAI 486

```



|    |     |   |     |
|----|-----|---|-----|
| Db | 427 | DGTHGAFDLSMRCISLPMNVIMAPSDENECRQMLYGVHNGSPAAYRPRGNSTSAVLE     | 486 |
| Qy | 487 | PTLTALTEICKAEVFRHHGSRILALLAMGSAWTPAVEGKXGLATVYNNRVRPFDAULTL   | 546 |
| Db | 487 | P-LEMLPTIGGVLRREERKILALICGTLLAQQLAAENLNATLVMMRVRKLDIELVLEM    | 545 |
| Qy | 547 | ARTHDVYVVEENVIVAGGAGSAINTFLOAKQILMPVCNIGLPRDFVEGSSREELISVGL   | 606 |
| Db | 546 | AAKHQVLVTVVEENAVIVAGGAGSGVNELLMKRSWVPLVNIIGLPDLFVPGQDEKRSBELG | 605 |
| Qy | 607 | DSKGILATITEQCA  | 620 |
| Db | 606 | DAAGIQRIEAMLA   | 619 |

## RESULT 4

AF0554  
1-deoxyxylulose-5-phosphate synthase [imported] - *Salmonella enterica* subsp. *enterica* serov. *typhimurium*  
C:Species: *Salmonella enterica* subsp. *enterica* serovar *typhi*  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AF0554  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov. *Paratyphi*  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AF0554  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-620 <PAR>  
A:Cross-references: GB:A513382; EIDN:CAD08878.1; PID:916501690; GSPDB:GN00176  
C:Genetics:  
A:Gene: dxs  
A:Superfamily: hypothetical protein C2814

|    |     |  |     |
|----|-----|--|-----|
| Db | 427 | DCQTHQAGDELYLRCIPDMVIMTDSDENECRQMLFTGYHNDGPAIVAVYPRNAGQVAL   | 486 |
| Qy | 486 | DEPLTLALTEGKAEVRHHSRIILAMGSMVTAVEAGKQLGATVVMRRPVKPPDQALVLE   | 545 |
| Db | 487 | 7P-LEKLPFGKGIKVRKHGEKAILNFGILIMPEAKVAEALNLTALVMRFKPPDQDLILE  | 545 |
| Qy | 546 | LRTIDVFLVTEENVYAGGASALNITFLQAOVKLMPVCHIGLPDFRVEGDSREELISLVG  | 605 |
| Db | 546 | MAQHDVALTLLEBNIMIGGAGSGVNEVLMAHRKPVVLNIGLPDFEIPQGTQSEAPAEISG | 605 |
| Qy | 606 | LDKGLIATITQFCA   | 620 |
| Db | 606 | LDAGIEAKIKAMLA   | 620 |

## RESULT 5

D64771  
 dds protein - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C:Accession: D64771  
 R:Latanez, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
 .A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426517; PMID:9278503  
 A:Accession: D64771  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-620 <BLAT>  
 A:Cross-references: GB:A0000148; GB:U00096; NID:g1786614; PIDN:AACT3523.1; PID:g1786622;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: dds  
 C:superfamily: hypothetical protein C2814

Query Match 62.2%; Score 2000.5; DB 2; Length 620;  
 Best Local Similarity 60.2%; Pred. No. 1,8e-132;  
 Matches 370; Conservative 144; Mismatches 128; Indels 3; Gaps 3;

Db 487 TP-LKLPFGKGVKRGKSKALINFGTILMPBAKVASINATIVDMRVKRLDLETLIE 545  
 Oy 546 LARHDVFTVYBNVNIAGAGSAINTEFLQOKVLMPCNIGLPDRFVQGSREBELSLVG 605  
 Db 546 MAASHBALVTVEENALMGAGSGVNEVLMARHKKPVPLNIGLPDFPIPGQTOEMRAELG 605  
 Oy 606 LDSKGLIATIEQCA 620  
 Db 606 LDAAGMEAKIKAMIA 620

## RESULT 6

B90688  
 1-deoxy-D-xylose 5-phosphate synthase [imported] - Escherichia coli (strain O157:H7, s  
 C.Species: Escherichia coli  
 C.Date: 19-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C.Accession: B90688  
 R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
 DNA Res. 8, 11-22, 2001  
 A.Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A.Reference number: A99629; UID:21156231; PMID:11258796  
 A.Accession: B90688  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-620 <HAY>  
 A.Cross-references: GB:BA000007; PIDN:BA03897.1; PID:G13359931; GSPDB:GN00154  
 A.Experimental source: strain O157:H7, substrain RIMD 0509552  
 C.Genetics:  
 A:Gene: BCS0474  
 C:Superfamily: hypothetical protein C2814

Query Match 62.0%; Score 1995.5; DB 2; Length 620;  
 Best Local Similarity 59.8%; Pred. No. 4e-132;  
 Matches 368; Conservative 115; Mismatches 129; Indels 3; Gaps 3;

Oy 7 YPLKNIHTPADIRASKQLOQLADENVGTLITHTVSIQGHFAAGLGTVALHYV 66  
 Db 8 YPLALVSTQELRLPKESLPKLCDELRLYLDVSRSQGHFASGLGTVALHYV 67  
 Oy 67 NTPVDQVMDVGHQAYPHKILTGKEMPTIRLIGVSAFPARDBSEYDAFGVGHSSSTSI 126  
 Db 68 NTPDQLIMDVGHQAYPHKILTGKEMPTIRLIGVSAFPARDBSEYDAFGVGHSSSTSI 127  
 Oy 68 NTPDQLIMDVGHQAYPHKILTGKEMPTIRLIGVSAFPARDBSEYDAFGVGHSSSTSI 127  
 Db 127 SAALGMAIASQLRGEDKKNVAIIIGDSITGMAVYEMNHAGDVNANLVLINDNMSISP 186  
 Oy 128 SAGIGIAVAAEKKGKRRVTCVIGDAITAGMAFEAMNHAGDIRPDLVLINDNMSISP 187  
 Db 128 SAGIGIAVAAEKKGKRRVTCVIGDAITAGMAFEAMNHAGDIRPDLVLINDNMSISP 187  
 Oy 187 PVGAMNNYITKVLSSKFYSVRESKALAKMPSVVELAKTEEHVGMIVPGTLFEEIG 246  
 Db 188 NGALNNHIAQLISGLYSLSLRGCKKVSVPPIKELKREHNIKGVVPGTLFEEIG 247  
 Oy 247 FNYFGPIDGHDVEMLVSTLENLKDLPVFLHVVTKKGKGYAPAKADPLAHGVPAEDPT 306  
 Db 248 FNYIGVDGHDVGLITLTKNMRLDKGPFLHMTKKGKGYAPAKADPTTFHVAVKFDS 307  
 Oy 248 FNYIGVDGHDVGLITLTKNMRLDKGPFLHMTKKGKGYAPAKADPTTFHVAVKFDS 307  
 Db 307 KQFLPKAASPHPTTEVFGKMLCMAAODERLIGTPAMREGSGLVESQKFPKRYPDV 366  
 Oy 308 SGCLPKSS-GGLPYSKIFEDMLCETAAKDKMAIITPAREEGGMBFSKFPDRYFDV 366  
 Db 308 SGCLPKSS-GGLPYSKIFEDMLCETAAKDKMAIITPAREEGGMBFSKFPDRYFDV 366  
 Oy 367 AIAEOHVAITLAAAGACGAKPVVAIYSTFLORGVDLIHVALONLMDLPAIDRAGLVGP 426  
 Db 367 AIAEOHVAITLAAAGACGAKPVVAIYSTFLORGVDLIHVALONLMDLPAIDRAGLVGP 426  
 Oy 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Db 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Oy 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Db 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Oy 486 DPTLTALEIGKAEVRHSGRIALIAAGSNVTPAVEAGKOLGATVNNRFPVFPDQALVE 545  
 Db 487 TP-LKLPFGKGVKRGKSKALINFGTILMPBAKVASINATIVDMRVKRLDLETLIE 545

Oy 546 LARHDVFTVYBNVNIAGAGSAINTEFLQOKVLMPCNIGLPDRFVQGSREBELSLVG 605  
 Db 546 MAASHBALVTVEENALMGAGSGVNEVLMARHKKPVPLNIGLPDFPIPGQTOEMRAELG 605  
 Oy 606 LDSKGLIATIEQCA 620  
 Db 606 LDAAGMEAKIKAMIA 620

## RESULT 7

F85338  
 1-deoxy-D-xylose 5-phosphate synthase [imported] - Escherichia coli (strain O157:H7, s  
 C.Species: Escherichia coli  
 C.Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 27-Nov-2001  
 C.Accession: F85338  
 R.Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.O.; Davis, N.W.; Lim, A.; Dialante, E.; Potamousis, K.; Apodaca,  
 Nature 409, 525-533, 2001  
 A.Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A.Reference number: A85480; UID:21074935; PMID:11206551  
 A.Accession: F85338  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-620 <STO>  
 A.Cross-references: GB:AE005174; UID:G12513276; PIDN:AG54770.1; GSPDB:GN00145; UWGP:205  
 A.Experimental source: strain O157:H7, substrain SDU933  
 C.Genetics:  
 A:Gene: dks  
 C:Superfamily: hypothetical protein C2814

Query Match 62.0%; Score 1995.5; DB 2; Length 620;  
 Best Local Similarity 59.8%; Pred. No. 4e-132;  
 Matches 368; Conservative 115; Mismatches 129; Indels 3; Gaps 3;

Oy 7 YPLKNIHTPADIRASKQLOQLADENVGTLITHTVSIQGHFAAGLGTVALHYV 66  
 Db 8 YPLALVSTQELRLPKESLPKLCDELRLYLDVSRSQGHFASGLGTVALHYV 67  
 Oy 67 NTPVDQVMDVGHQAYPHKILTGKEMPTIRLIGVSAFPARDBSEYDAFGVGHSSSTSI 126  
 Db 68 NTPDQLIMDVGHQAYPHKILTGKEMPTIRLIGVSAFPARDBSEYDAFGVGHSSSTSI 127  
 Oy 68 NTPDQLIMDVGHQAYPHKILTGKEMPTIRLIGVSAFPARDBSEYDAFGVGHSSSTSI 127  
 Db 127 SAALGMAIASQLRGEDKKNVAIIIGDSITGMAVYEMNHAGDVNANLVLINDNMSISP 186  
 Oy 128 SAGIGIAVAAEKKGKRRVTCVIGDAITAGMAFEAMNHAGDIRPDLVLINDNMSISP 187  
 Db 128 SAGIGIAVAAEKKGKRRVTCVIGDAITAGMAFEAMNHAGDIRPDLVLINDNMSISP 187  
 Oy 187 PVGAMNNYITKVLSSKFYSVRESKALAKMPSVVELAKTEEHVGMIVPGTLFEEIG 246  
 Db 188 NGALNNHIAQLISGLYSLSLRGCKKVSVPPIKELKREHNIKGVVPGTLFEEIG 247  
 Oy 247 FNYFGPIDGHDVEMLVSTLENLKDLPVFLHVVTKKGKGYAPAKADPLAHGVPAEDPT 306  
 Db 248 FNYIGVDGHDVGLITLTKNMRLDKGPFLHMTKKGKGYAPAKADPTTFHVAVKFDS 307  
 Oy 248 FNYIGVDGHDVGLITLTKNMRLDKGPFLHMTKKGKGYAPAKADPTTFHVAVKFDS 307  
 Db 307 KQFLPKAASPHPTTEVFGKMLCMAAODERLIGTPAMREGSGLVESQKFPKRYPDV 366  
 Oy 308 SGCLPKSS-GGLPYSKIFEDMLCETAAKDKMAIITPAREEGGMBFSKFPDRYFDV 366  
 Db 308 SGCLPKSS-GGLPYSKIFEDMLCETAAKDKMAIITPAREEGGMBFSKFPDRYFDV 366  
 Oy 367 AIAEOHVAITLAAAGACGAKPVVAIYSTFLORGVDLIHVALONLMDLPAIDRAGLVGP 426  
 Db 367 AIAEOHVAITLAAAGACGAKPVVAIYSTFLORGVDLIHVALONLMDLPAIDRAGLVGP 426  
 Oy 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Db 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Oy 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Db 427 DGPTHAGADYSYKRCIPNMLIAPADENECROMLTGFOH-GRASVYPRGKPGAAI 485  
 Oy 486 DPTLTALEIGKAEVRHSGRIALIAAGSNVTPAVEAGKOLGATVNNRFPVFPDQALVE 545  
 Db 487 TP-LKLPFGKGVKRGKSKALINFGTILMPBAKVASINATIVDMRVKRLDLETLIE 545  
 Oy 546 LARHDVFTVYBNVNIAGAGSAINTEFLQOKVLMPCNIGLPDRFVQGSREBELSLVG 605  
 Db 546 MAASHBALVTVEENALMGAGSGVNEVLMARHKKPVPLNIGLPDFPIPGQTOEMRAELG 605

QY 606 LDSKILATIEQFCA 620  
 DB 606 LDAGGAKIKAMLA 620

## RESULT 8

deoxyxylulose-5-phosphate synthase XF2249 (imported) - Xylella fastidiosa (strain 9asc)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: A82582

Ranonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A82582

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-670 <SIM>

A:Cross-References: GB:AEO04037, GB:AEO03849, NID:g9107394, PIDN:AAP85048.1; GSPDB:GN001

A:Experimental source: strain 9asc

R:Simson, A.J.G.; Relinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carreir, D.M.; Carreir, H

as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.B.; Kurmae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, B

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak

M.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2249

C:Superfamily: hypothetical protein C2814

## Query Match

Best Local Similarity 59.3%; Score 1951; DB 2; Length 670;

Matches 365; Conservative 98; Mismatches 147; Indels 6; Gaps 3;

QY 4 TDDYPLKNIHTPADRALSKDLOOLADEVGYLTHTVYISGHPAAGLTVETLVAH 63  
 DB 39 STCYPRLSIQIPEDLRTQSESLAVADHLYLIESVLSGHPAAGLTVETLVAH 98  
 QY 64 YFNTPDVQVMDVGHQAVPHKILTGKRMFTIRTLGVSAPPARDESEYDAFVGHSS 123  
 DB 99 YVCTPDIHVDVGHQAVPHKILTGKRMFTIRTLGVSAPPARDESEYDAFVGHSS 158  
 QY 124 TGISALGAIAISQLEGEDEKRVVAIIIGDSITGMAVEAMNHAGVNV--ANLTVLIND 181  
 DB 159 TGISALGAIAISQLEGEDEKRVVAIIIGDSITGMAVEAMNHAGVNV--ANLTVLIND 218  
 QY 182 MSISPPVGMANNVLTKVLSKFEYSVRESKKA--KAPSVWELAKTEEHVKMIVPG 239  
 DB 219 MSISAVGGLTVMGLATGSKALNMRGSKRIFGDKTNATAPRLRREEMKMPVPS 278  
 QY 240 TLFEEGPNVFPPIGDHVEMLVSTLENKDLTGVPFLHVTYTKKGYAPAEKDLAHHG 299  
 DB 279 TLFEEGPNVFPPIGDHVEMLVSTLENKDLTGVPFLHVTYTKKGYAPAEKDLAHHG 338  
 QY 300 VPAPFDKPLPKAAPSHPFTYTVFGRWLCDMAADERLLGTPAMEGSGLVFESQKF 359  
 DB 339 VAPFPDQGLIYAGAKK--QYTVDFSEMLCDMAAVERPLTATPAMEGSGLVFESQEX 396  
 QY 360 PRRYVDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQMLDPLD 419  
 DB 397 PRRYVDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQMLDPLD 456

QY 420 PAILVGGDPGTHAGAFDYSTYKRCIPNMLIAPADENECROKLTGPOHGPASVYPRGK 479  
 DB 457 RGVVGPBGATHAQNLDSFLRCVPMMLIAPADENECROKLTGPOHGPASVYPRGK 516  
 QY 480 GPGADIDPFLTALISGRKRVNHSRIATILAMGSMVPAVAGKQIAGTVNMFPVKPD 539  
 DB 517 GPGVPSASLKVLTGVGAQIARHSSTRITALLGFGVCVAPAEQVGRRLGLTVNMFIRPD 576  
 QY 540 QALVELARTHDVFTVYEVNVIAGAGSALNTFLQAKVLMVPCNIGLPDRFVQSGR 599  
 DB 577 RTLLLEARTHEGVTTEEDNVVAGAGSAGVALLNARGLVPIVHGLPDAFQHASRD 636  
 QY 600 LLSLVGDSKILATI 615  
 DB 637 LLAAGIDMAGVVAL 652

## RESULT 9

B64172

dse protein - Haemophilus influenzae (strain Rd KX20)

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 24-Sep-1999

C:Accession: B64172

R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, B.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:9530630; PMID:7542800

A:Accession: B64172

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-625 <TIGR>

A:Cross-References: GB:UJ2822; GB:L42023; NID:g1574265; PIDN:AAC23088.1; PID:g1574278; T

A:Contents: dse

C:Genetics:

A:Gene: dse

C:Superfamily: hypothetical protein C2814

## Query Match

Best Local Similarity 59.6%; Score 1937.5; DB 2; Length 625;

Matches 368; Conservative 102; Mismatches 142; Indels 5; Gaps 3;

QY 6 DYPILKNIHTPADRALSKDLOOLADEVGYLTHTVYISGHPAAGLTVETLVAH 65  
 DB 7 NYPLILINSPEDELRILNKQOLPOLCELRAYLTLESVQSGHLAGSGLTVETLVAH 66  
 QY 66 FNTFVDQVMDVGHQAVPHKILTGKRMFTIRTLGVSAPPARDESEYDAFVGHSS 125  
 DB 67 YKTPFDQVMDVGHQAVPHKILTGKRMFTIRTLGVSAPPARDESEYDAFVGHSS 126  
 QY 126 ISALGAIAISQLEGEDEKRVVAIIIGDSITGMAVEAMNHAGVNV--ANLTVLIND 185  
 DB 127 ISALGAIAISQLEGEDEKRVVAIIIGDSITGMAVEAMNHAGVNV--ANLTVLIND 186  
 QY 186 PPGVGMANNVLTKVLSKFEYSVRESKKA--KAPSVWELAKTEEHVKMIVPG 243  
 DB 187 ENVALINNHARIPSGLSYSTRIDGSKKILDKVPIPNPKKTEEHVKMIVPG 246  
 QY 244 ELGPNVFPPIGDHVEMLVSTLENKDLTGVPFLHVTYTKKGYAPAEKDLAHHG 303  
 DB 247 ELGPNVFPPIGDHVEMLVSTLENKDLTGVPFLHVTYTKKGYAPAEKDLAHHG 306  
 QY 304 DPTDFPLPKAAPSHPFTYTVFGRWLCDMAADERLLGTPAMEGSGLVFESQKF 363  
 DB 307 DPTDFPLPKAAPSHPFTYTVFGRWLCDMAADERLLGTPAMEGSGLVFESQKF 364  
 QY 364 PDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQMLDPLD 423  
 DB 365 PDVAIAEGHATVLAAGACGAKPVVAIYSTFLQSGYQLHVDVALQMLDPLD 424  
 QY 424 VGGPDPGTHAGAFDYSTYKRCIPNMLIAPADENECROKLTGPOHGPASVYPRGK 483

Db 425 VGADGATHQGAADISFMRGCIPIIMLTBPSDBNCHQMLYTGCGKPAAYRPGNAVG 484  
 Oy AIDPTLTALBEIGAEVREHSGRIALLANGSMVTPAVEAGTGVMMRFPVCPDQALV 543  
 Db 445 KLTLP-LEMLPICKSLIKKQKALINFGTLLPSALBESKLNATVMMRFPVCPDQALV 543  
 Oy 544 LELANTADVPTVEENVVAGAGSAINFLQAKVLMPCVNGIDPFRVGGSSREBLISL 603  
 Db 544 NVLAQTHYLVTLBERALQGGAGSAVAEVLNSGKSTALLQGLPDPYFIPQATQCEALAD 603  
 Oy 604 VGIDSKGILATEGPCA 620  
 Db 604 LGDTKGIEKTLNFIH 620

## RESULT 10

B81978

Probable 1-deoxyxylulose-5-phosphate synthase NMA0589 [imported] - *Neisseria meningitidis*  
 C/Species: *Neisseria meningitidis*  
 C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C/Accession: B81978

R.Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
 J.; Holtz, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
 Nature 404, 502-506, 2000

A>Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
 A/Reference number: A81775; MUID:20222556; PMID:10761919

A/Accession: B81978  
 A/Status: preliminary

A/Molecule type: DNA  
 A/Residues: 1-637 <PAR>

A/Cross-references: GB:AL162753; GB:AL157959; NID:G7379120; PIDN:CA883880.1; PID:G737932  
 A/Experimental source: serogroup A, strain Z2491

C/Genetics:

A/Gene: OX5, NMA0589

C/Superfamily: hypothetical protein C2814

Query Match 55.2%; Score 1776; DB 2; Length 637;  
 Best Local Similarity 55.3%; Pred. No. 1.1e-116;  
 Matches 347; Conservative 92; Mismatches 155; Indels 34; Gaps 5;

Oy 8 PLKNIHTPADIRALSKDQQLADBEVRYLTHTVSISSGHFAAGLSTVELTVALHYFN 67  
 Db 5 PLDLIDSPQDLRLDKQPLRAGELRTLLSVGGTGHFASNLGAVELTVALHYFN 64  
 Oy 68 TPVDOLVMDVGEQAYPHKILTGKERMPTITLGGVSAFPARDESEYDAFGVGSSTIS 127  
 Db 65 TPEDKLVMDVGEQAYPHKILTGKERMPTITLGGVSAFPARDESEYDAFGVGSSTIS 124  
 Oy 128 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 187  
 Db 125 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 184  
 Oy 188 VGAMNNYLTLYLSKRY---SSVRESKALAKMPSVMEIARKE-----EHVKG 234  
 Db 185 VGALPKYLAASNVVDMGILLSTVAKGTGKYLDKIPGAMBERAKQVEHKIKTLAEERHAKQ 244  
 Oy 235 MIYVGTLEFELGFPYRPGIDGDEMVLSTLENLKDITGSPVFLHVVTKKGYAPAEKDP 294  
 Db 245 SL---SLFENFGFRYTGPDGHNENLVLDLADSKSKGQQLHVTIKKNGYGLAENDP 301  
 Oy 295 LAYHGPAPDPTKDLFLPKAA-----PSHPYTEVFGMLCDMAQDERLLGIT 343  
 Db 302 VKYHAYAN-----LPKESAQMPSSEKPKPAKFTTYQVFGKMLCDRAADSLVAIT 354  
 Oy 344 PAMRGGSLVGFSGKFPNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQRYDOL 403  
 Db 355 PAMRGGSLVGFSGKFPNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQRYDOL 414  
 Oy 404 IHDAIQLNLTDMFLADRAIGVDPGPTTHAGADYSYMRGCIPIIMLTBPSDBNCHQMLY 463  
 Db 415 VHDIALQNLPLVFAVNDAGIVGADGPTTHAGYDLSPLRCPVNMIVAAPSDBNCHQMLY 474  
 Oy 464 GFOHNGPASVRYPRGKGPAAIDPTLTALBEIGAEVREHSGRIALLANGSMVTPAVEAGK 523

Db 475 CQADAPAAVRYRGGTGPVSDGMEVIEIGGIIIRRECKTAFLAFSGMVAPALAVAG 534  
 Oy 524 QLGATVMMRFPVCPDQALVLEARTDVVVEENVVAGAGSAINFLQAKVLMPCV 583  
 Db 535 KLMRTVADNRFPVCPDQALVLEARTDVVVEENVVAGAGSAINFLQAKVLMPCV 583  
 Oy 584 NIGLPDRFVGGSSREBLISL 611  
 Db 595 LIGVADVTGTHGDPKRLDLDGLSABAV 622

## RESULT 11

D81034

1-deoxyxylulose-5-phosphate synthase NMB1867 [imported] - *Neisseria meningitidis* (strain

C/Species: *Neisseria meningitidis*  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C/Accession: D81034

R.Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.  
 Science 287, 1809-1815, 2000

A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A/Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

A/Reference number: A81000; MUID:20175755; PMID:10710307

A/Accession: D81034  
 A/Status: preliminary

A/Molecule type: DNA  
 A/Residues: 1-637 <TGT>

A/Cross-references: GB:AE002536; GB:AE002098; NID:G7227115; PIDN:AAE42201.1; PID:G722712  
 A/Experimental source: serogroup B, strain MC58

C/Genetics:

A/Gene: NMB1867

C/Superfamily: hypothetical protein C2814

Query Match 55.2%; Score 1775; DB 2; Length 637;  
 Best Local Similarity 54.8%; Pred. No. 1.2e-116;  
 Matches 344; Conservative 97; Mismatches 153; Indels 34; Gaps 5;

Oy 8 PLKNIHTPADIRALSKDQQLADBEVRYLTHTVSISSGHFAAGLSTVELTVALHYFN 67  
 Db 5 PLDLIDSPQDLRLDKQPLRAGELRTLLSVGGTGHFASNLGAVELTVALHYFN 64  
 Oy 68 TPVDOLVMDVGEQAYPHKILTGKERMPTITLGGVSAFPARDESEYDAFGVGSSTIS 127  
 Db 65 TPEDKLVMDVGEQAYPHKILTGKERMPTITLGGVSAFPARDESEYDAFGVGSSTIS 124  
 Oy 128 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 187  
 Db 125 AALGMAIASQURGEDKXKVAIIGDSITGMAAYEAMNAGDVANLVLINDNMSISPP 184  
 Oy 188 VGAMNNYLTLYLSKRY---SSVRESKALAKMPSVMEIARKE-----EHVKG 234  
 Db 185 VGALPKYLAASNVVDMGILLSTVAKGTGKYLDKIPGAMBERAKQVEHKIKTLAEERHAKQ 244  
 Oy 235 MIYVGTLEFELGFPYRPGIDGDEMVLSTLENLKDITGSPVFLHVVTKKGYAPAEKDP 294  
 Db 245 SL---SLFENFGFRYTGPDGHNENLVLDLADSKSKGQQLHVTIKKNGYGLAENDP 301  
 Oy 295 LAYHGPAPDPTKDLFLPKAA-----PSHPYTEVFGMLCDMAQDERLLGIT 343  
 Db 302 VKYHAYAN-----LPKESAQMPSSEKPKPAKFTTYQVFGKMLCDRAADSLVAIT 354  
 Oy 344 PAMRGGSLVGFSGKFPNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQRYDOL 403  
 Db 355 PAMRGGSLVGFSGKFPNRYFDVAIAEOHAVTLAAGACGAKPVVAIYSTFLQRYDOL 414  
 Oy 404 IHDAIQLNLTDMFLADRAIGVDPGPTTHAGADYSYMRGCIPIIMLTBPSDBNCHQMLY 463  
 Db 415 VHDIALQNLPLVFAVNDAGIVGADGPTTHAGYDLSPLRCPVNMIVAAPSDBNCHQMLY 474  
 Oy 464 GFOHNGPASVRYPRGKGPAAIDPTLTALBEIGAEVREHSGRIALLANGSMVTPAVEAGK 523

Db 475 CYQADAPAAVRYPRGTGCAFPVSDGMEYVEIKGIIIRREGKTAFLAGSNVAPLANAE 534  
 QY 524 QLGATVVMNRFPVKPDPQALVLEIARTHDVFTVEENVVAGAGSAINTFLQAKYLMKMEVC 583  
 Db 535 KLNATVADMRFVKPILDBELIVLARSHDRIVTLENAEGGAGAVLEIARHGICKEVLT 594  
 QY 584 NIGLPDRPVEQSSREELSLVGLDSKGI 611  
 Db 595 LLGVADVTYGRKDPKLLDDIGLSABAV 622

## RESULT 12

das protein [imported] - Buchnera sp. (strain APS)

C.Species: Buchnera sp.  
 C.Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C.Accession: AB4984  
 R.Shigenobu, S.; Matanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
 Nature 407, 81-86, 2000  
 A.Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.  
 A.Reference number: AB4930; MUID:20445173; PMID:10939077  
 A.Accession: AB4984  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-608 <STO>  
 A.Cross-references: GB:AP000398; GSPDB:GN00144  
 A.Experimental source: strain APS  
 C.Genetics:  
 A.Gene: dks; BU464  
 C.Superfamily: hypothetical protein C2814

Query Match 51.8%; Score 1667; DB 2; Length 608;  
 Best Local Similarity 52.8%; Pred. No. 4.4e-109;  
 Matches 325; Conservative 118; Mismatches 146; Indels 26; Gaps 7;

QY 7 YPLKNHTPADIRALSKDQLOLADVEVGYTFTVTSISGGHPAAGLTVETVLAHYF 66  
 Db 8 YPILSPANSVENIRLISYEOLPQLCFELREYLLDVVISISKHFPAGLGVETVLAHYV 67  
 QY 67 NTPVDOLVMDVGHQAYPHKILNGRKEKMPITRTLGVSAPARDESBYDAGVGHSTSI 126  
 Db 68 NTPEDNLMDGCHQAYPHKILNGRKEKINSIRKNGHLSFPRESSEYDSLVGHSTSI 127  
 QY 127 SAALGMALASQLRGDKKXVALIGDSITGCMAYEAMNHAGDVANLVLINDNDMSIS 186  
 Db 128 SAGLGMASIAEKEGKRNKTCIIGGAMTAGMAFALINAGSIOENLVLINDNDMSISR 187  
 QY 187 PVGAANNVLTKLSS-----KFSYSVREESKKALKMSVWEIAKTEEHYKGMIVPGL 241  
 Db 188 NVGALNKLH-KILRSVQNTQKRRKIRLNLKLFK-----DKRIQNH--SISFNSI 236  
 QY 242 FEELGFNYFGPIDGHDVEMVSTLENKDLTGVPVLAHVYTKKGYAPAEKDPVLAHGP 301  
 Db 227 FSNIGCKVLGPPDGNITSIINTLKIKKNGKTYLLHLYTKKGYLPALNPIKMTIS 296  
 QY 302 ADPPTKDFLPKAAEPHPHYTEVFGRLCDMAAOBERLLGITPARBESSGLVEBSQKEPN 361  
 Db 297 SRD-----SSVRSLSYSDFGTWLCZAIAPFKKLIATTPAACESGAWKFSRLPN 348  
 QY 362 RFVDVIAIEGHEAVTLAAGACQAGAPVAAYSTFLORGVDOLIHVALQNDMLFALDRA 421  
 Db 349 QYFDVIAIEGHEAVTLAAGLAIISGYFVVISYTFPQRAYDOLIHVALQNKLSVLPADRA 408  
 QY 422 GLVPGDGPTHAGAFYSYTRCIPNMLIMAPADENECRQMLTTGFQHH-GRPASVRYRQKG 480  
 Db 409 GIVGNDGQTHGQVFLAYIRCIPIGIVIMTPSNMENCROMLYTGVMENKGPSPVRYRQKG 468  
 QY 481 PGAAIDPTLTALIEGAEVHHGSRALILAMGSMTPVAVBAKGLGATVVMNRFPVKPDPQ 540  
 Db 469 VQELLNP-MNRPIPKSLIKRQKILAILNPGILHLNAYCAEKDALVDMRFVKPILDK 527  
 QY 541 ALVLEIARTHDVFTVEENVVAGAGSAINTFLQAKYLMKPMVNCIGLPDRPVEQSSREEL 600

Db 528 SMILKLSQNKPFITLBERGVISGAGSAVNEFIVNKLFLPVNLIGLPDPTIPQGTQEEI 587  
 QY 601 LSLVGLDSKGIAT 615  
 Db 588 RHVYKLDSEGIYKQI 602

## RESULT 13

A70376

conserved hypothetical protein ag\_881 - Aquifex aeolicus

C.Species: Aquifex aeolicus  
 C.Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 24-Sep-1999  
 C.Accession: A70376

R.Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ore  
 V.

Nature 392, 353-358, 1998

A.Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A.Reference number: A70300; MUID:98196666; PMID:9537320

A.Accession: A70376  
 A.Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA  
 A.Residues: 1-628 <AGP>

A.Cross-references: GB:AE00712; MID:92983411; PIDN:AAC07004.1; PID:G2983421; GB:AE00065;  
 A.Experimental source: strain VFS  
 C.Genetics:  
 A.Gene: ag\_881  
 C.Superfamily: hypothetical protein C2814

Query Match 50.4%; Score 1620.5; DB 2; Length 628;  
 Best Local Similarity 51.9%; Pred. No. 8.6e-106;  
 Matches 320; Conservative 105; Mismatches 174; Indels 17; Gaps 8;

QY 7 YPLKNHTPADIRALSKDQLOLADVEVGYTFTVTSISGGHPAAGLTVETVLAHYF 66  
 Db 5 YEIKQYKGPEDIKNYETLOKLAQEVROYIINVTSKNGHVPSPISGLVETVLAHYF 64  
 QY 67 NTPVDOLVMDVGHQAYPHKILNGRKEKMPITRTLGVSAPARDESBYDAGVGHSTSI 126  
 Db 65 NTPEDVIMDIGHQYPMKILTDKEQPTLRQYKGISGFRRBSIYDAFGAGHSTSI 124  
 QY 127 SAALGMALASQLRGDKKXVALIGDSITGCMAYEAMNHAGDVANLVLINDNDMSI 184  
 Db 125 SAALGFRIGDKLKEKEDYVLAIVGDALTRGMAVEALNNAGHTRPRFIYILNDMSI 184  
 QY 185 SPVGAANNVLTKLSSKFSYSVREESKKALAYN-PSWEIAKTEEHYKGMIVPGLFE 243  
 Db 186 SPVGAISTYLNRIISGFVQETROKIFLQHGRTPLRMKLTBEFLKGLISPGVIF 244  
 QY 244 ELGFNYFGPIDGHDVEMVSTLENKDLTGVPVLAHVYTKKGYAPAEKDPVLAHGP 303  
 Db 245 ELGFNYFGPIDGHDVEMVSTLENKDLTGVPVLAHVYTKKGYAPAEKDPVLAHGP 304  
 QY 304 D-PTKDFLPKAAEPHPHYTEVFGRLCDMAAOBERLLGITPARBESSGLVEBSQKEPN 362  
 Db 305 KYESGEIIRKSSP---PLTWSVFGKALVELABDEKIVALTIPARRSSGLVEFAKFPDR 361  
 QY 363 YEDVIAIEGHEAVTLAAGACQAGAPVAAYSTFLORGVDOLIHVALQNDMLFALDRA 422  
 Db 362 FFDVIAIEGHEAVTLAAGLAIISGYFVVISYTFPQRAYDOLIHVALQNKLSVLPADRA 421  
 QY 422 GLVPGDGPTHAGAFYSYTRCIPNMLIMAPADENECRQMLTTGFQHH-GRPASVRYRQKG 480  
 Db 481 GLVPGDGPTHAGAFYSYTRCIPNMLIMAPADENECRQMLTTGFQHH-GRPASVRYRQKG 480  
 QY 483 AAIDPT--LTALETGAEVHHGSRALILAMGSMTPVAVBAKGLGATVVMNRFPVKPDPQ 540  
 Db 482 V--PTGPFKKIIEGTVEELIGEDCVILAWGPVVALAAEKLVEGIRVGVNARFV 538  
 QY 536 KPFGDALVLEIARTHDVFTVEENVVAGAGSAINTFLQAKYLMKPMVNCIGLPDRPVEQSS 595  
 Db 539 KPNDBKRLDRLAKRYDPTFVEEDVTYVGGSGVLEFPAAEGIMKRYINLGVPRFIEHG 598  
 QY 596 SREELSLVGLDSKGI 611

Db 599 KODILNIVGIDABGI 614

## RESULT 14

AD3439 1-deoxyxylulose-5-phosphate synthase (EC 4.2.1.-) [imported] - *Brucella melitensis* (strain

C/Species: *Brucella melitensis*  
C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002

C/Accession: AD3439

R/Deleccio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*

A/Reference number: AD3252; PMID:1175668

A/Accession: AD3439

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-643 <KUR>

A/Cross-references: GB:AE008917; PIDN:AA152679.1; PID:q17983505; GSPDB:GN00190

A/Experimental source: strain 16M

A/Genetics:

A/Map position: 1

C/Superfamily: hypothetical protein C2814

C/Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 49.5%; Score 1591.5; DB 2; Length 643;

Best Local Similarity 51.4%; Pred. No. 9, 6e-104;

Matches 318; Conservative 106; Mismatches 182; Indels 13; Gaps 7;

```

QY 8 PLKNIHTPADIALSKDQLQDLADVRGTLTTSISGHPAAGLGVETLVAALHYEN 67
DB 7 PLDDKPTDRLRLAEQDLPLAEELRTELIDAVSTGGHAGLGVETLVAALHYEN 66
QY 68 TPVDQLVMDVGHQAYPHKILITGKEMPTIRITLQVSAPPADESEYDAFGVGHSTIS 127
DB 67 TPYDRIMVGHQAYPHKILITGKEMPTIRITLQVSAPPADESEYDAFGVGHSTIS 126
QY 128 AALGMAIASQLRGEDKMAIIGDGSITGMAVEANMHADVNAVNLVILINDNISISP 187
DB 127 AGGMAVASLSEKKNVILVIGDGSMAKAYEANNAGALDARLIVLINDNISIAFP 186
QY 188 VGAMNNYLVKLSKTVSSVRESKALAKMP5-VNELAKTBEHYKGMIVPGLFEEELG 246
DB 187 TGMASVLAIVSGRTYRSVBAKQVAGL.PKFLQDKARKSEVAPAFPGTILFEEELG 246
QY 247 FNYFGSIDGHVEMVSTLENLKD-LTGPVFLAVYTKKGGYAPAKDPLAHGVPAPFP 305
DB 246 FNYFGSIDGHVEMVSTLENLKD-LTGPVFLAVYTKKGGYAPAKDPLAHGVPAPFP 305
QY 306 FYVVGIDGHVEMVSTLENLKD-LTGPVFLAVYTKKGGYAPAKDPLAHGVPAPFP 306
DB 305 FYVVGIDGHVEMVSTLENLKD-LTGPVFLAVYTKKGGYAPAKDPLAHGVPAPFP 306
QY 306 TNDPLKAPSPHPTTYEVRGRLCDMAADDERLIGITPAMEGSGLVEFSQKFPNRYED 365
DB 307 ITGKQAK-PPANAPSTTKIFGTSLIEARHDXIVAVTAMPFGTGLDFEGLFPRVFD 365
QY 366 VALAEQAVTLAAGQACQAKPVAIYSTFLQRGYQLIHVVALQNLMDLPAIDRAGLVG 425
DB 365 VALAEQAVTLAAGQACQAKPVAIYSTFLQRGYQLIHVVALQNLMDLPAIDRAGLVG 425
QY 426 VGLAEQAVTLAAGQACQAKPVAIYSTFLQRGYQLIHVVALQNLMDLPAIDRAGLVG 425
DB 425 VGLAEQAVTLAAGQACQAKPVAIYSTFLQRGYQLIHVVALQNLMDLPAIDRAGLVG 425
QY 426 PDGPTAGAFDYSYMCIPMMLIMAPADENECQMLTTGFQ-HGASVRYPRGKPGAA 484
DB 425 PDGPTAGAFDYSYMCIPMMLIMAPADENECQMLTTGFQ-HGASVRYPRGKPGAA 484
QY 485 IDPTLALAIQKAEVHHSRIAILANGSVTPAFAAGQLGA-----TVNMRPYKPPD 539
DB 484 LPERGSLVLEIGKRIYREGTKVALLSFGTRLQECILAAAEELGALSTVAARAPKPID 545
QY 540 QALVLEARTHDVFTVEENVVINGAGSALNTFLQAKYL---MPVCNIGLPPRF 596
DB 546 HDLIRLAEHEVLYVVEBGA-V-GGFGSHVLOFLADGLDGRFKYRALTLPDIYDHRK 604
QY 597 REELSLVGLDCKGLIATI 615

```

Db 605 PDAMTAEAGLDRTGIVRTV 623

## RESULT 15

C87505 1-deoxyxylulose-5-phosphate synthase [imported] - *Caulobacter crescentus*

C/Species: *Caulobacter crescentus*  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C/Accession: C87505

R/Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Koon

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: A67249; PMID:1125967

A/Accession: C87505

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-640 <STO>

A/Cross-references: GB:AE005673; NID:G13423547; PIDN:AAK24039.1; GSPDB:GN00148

A/Genetics:

A/Map position: 1

C/Superfamily: hypothetical protein C2814

Query Match 47.4%; Score 1523.5; DB 2; Length 640;

Best Local Similarity 50.7%; Pred. No. 5, 7e-99;

Matches 315; Conservative 97; Mismatches 196; Indels 13; Gaps 7;

```

QY 3 LTTDPLKNIHTPADIALSKDQLQDLADVRGTLTTSISGHPAAGLGVETLVAAL 62
DB 1 MSSKPLDITLAPADTGLSLAEIKOLAAEVRAETIDAVSTGGHAGLGVETLVAAL 60
QY 63 HYVNTPVQDLVMDVGHQAYPHKILITGKEMPTIRITLQVSAPPADESEYDAFGVGH 122
DB 61 HNVETPDVYVMDVGHQAYPHKILITGKEMPTIRITLQVSAPPADESEYDAFGVGH 120
QY 123 STSISALGMAIASQLRGEDKMAIIGDGSITGMAVEANMHADVNAVNLVILINDNM 182
DB 121 ATSIAGLFCAPARAKEDNSVIAVIGDGSIGMAVEANNAATDTTKRLIVLINDNM 180
QY 183 SISPVGAMNNYLVKLSKTVSSVRESKALAKMP5-VNELAKTBEHYKGMIVPGLT 241
DB 181 SIAPVGSAMSAVLAIVSGRTYRSVBAKQVAGL.PKFLQDKARKSEVAPAFPGTIL 240
QY 242 FEEELGFTYFGSIDGHVEMVSTLENLKD-LTGPVFLAVYTKKGGYAPAKDPLAHG 300
DB 241 FEEELGFTYFGSIDGHVEMVSTLENLKD-LTGPVFLAVYTKKGGYAPAKDPLAHG 300
QY 301 PADEPTKDFLKAAPSPHPTTYEVRGRLCDMAADDERLIGITPAMEGSGLVEFSQ 360
DB 300 VKEFDVVTGQQQKAAAGP-PSYTKVFADELINQAEKDXIVAVTAMPFGTGLDFEGL 359
QY 361 NRYFPVLAIEQAVTLAAGQACQAKPVAIYSTFLQRGYQLIHVVALQNLMDLPAID 420
DB 360 ERTPDVGLAEQAVTLAAGQACQAKPVAIYSTFLQRGYQLIHVVALQNLMDLPAID 419
QY 421 AGLVGPDGPTAGAFDYSYMCIPMMLIMAPADENECQMLTTGFQ-HGASVRYPRG 479
DB 420 AGLVGPDGPTAGAFDYSYMCIPMMLIMAPADENECQMLTTGFQ-HGASVRYPRG 479
QY 480 GPGALIDPTLALAIQKAEVHHSRIAILANGSVTPAFAAGQLGA-----AGQ 534
DB 480 GPGALIDPTLALAIQKAEVHHSRIAILANGSVTPAFAAGQLGA-----AGQ 534
QY 535 VKPFDQALVLELARTHDVFTVEENVVINGAGSALNTFLQAKYL---MPVCNIGL 591
DB 540 AKPLDLDLILRLAEHEVLYVVEBGA-V-GGFGSHVLOFLADGLDGRFKYRALTLP 598
QY 592 VEGSREELSLVGLDCKGLI 612
DB 599 QDQKFPDAMTAEAGLDRTGIVRTV 619

```

Thu Jan 29 17:41:29 2004

us-09-941-947a-6.rpt

Page 9

Search completed: January 29, 2004, 15:55:29  
Job time : 20.2424 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 11.6752 Seconds

(without alignment)  
2497.314 Million cell updates/sec

Title: US-09-941-947a-6

Sequence: 1 MCLTIDYPLAKNHTPADIR.....LSLVGLDSKGLATTEQPCA 620

Scoring table: BLOSUM62  
Gapop 10.0 , Gapect 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1          | 2114   | 65.7        | 621    | 1  | DXS_VIBVU   |
| 2          | 2097.5 | 65.2        | 626    | 1  | DXS_VIBCH   |
| 3          | 2069.5 | 64.4        | 627    | 1  | DXS_PSEAE   |
| 4          | 2042   | 63.5        | 622    | 1  | DXS_SHEON   |
| 5          | 2015   | 62.7        | 619    | 1  | DXS_YERBP   |
| 6          | 2014.5 | 62.6        | 619    | 1  | DXS_SALTI   |
| 7          | 2013.5 | 62.6        | 619    | 1  | DXS_SALTY   |
| 8          | 2008.5 | 62.5        | 636    | 1  | DXS_RALSO   |
| 9          | 2000.5 | 62.2        | 619    | 1  | DXS_ECOLI   |
| 10         | 1999.5 | 62.2        | 619    | 1  | DXS_ECOL6   |
| 11         | 1995.5 | 61.9        | 619    | 1  | DXS_ECOL5   |
| 12         | 1989.5 | 61.8        | 638    | 1  | DXS_XANCP   |
| 13         | 1986.5 | 61.8        | 638    | 1  | DXS_XANCP   |
| 14         | 1969   | 61.2        | 614    | 1  | DXS_PASMU   |
| 15         | 1951   | 60.7        | 635    | 1  | DXS_XYFPA   |
| 16         | 1937.5 | 60.2        | 635    | 1  | DXS_HAETN   |
| 17         | 1795   | 55.8        | 626    | 1  | DXS_WICBR   |
| 18         | 1776   | 55.2        | 637    | 1  | DXS_NEIMA   |
| 19         | 1775   | 55.2        | 637    | 1  | DXS_NEIMA   |
| 20         | 1667   | 51.8        | 608    | 1  | DXS_BUCNI   |
| 21         | 1629.5 | 50.7        | 585    | 1  | DXS_BUCAP   |
| 22         | 1620.5 | 50.4        | 628    | 1  | DXS_AQUAB   |
| 23         | 1592.5 | 49.5        | 643    | 1  | DXS_BRUSU   |
| 24         | 1591.5 | 49.5        | 643    | 1  | DXS_BRUSU   |
| 25         | 1543.5 | 48.0        | 637    | 1  | DXS_RHILC   |
| 26         | 1523.5 | 47.4        | 640    | 1  | DXS_RHILC   |
| 27         | 1516.5 | 47.2        | 622    | 1  | DXS_THETN   |
| 28         | 1499.5 | 46.6        | 639    | 1  | DXS_AGRTS   |
| 29         | 1498.5 | 46.6        | 639    | 1  | DXS_AGRTS   |
| 30         | 1460.5 | 45.4        | 645    | 1  | DXS_RHIME   |
| 31         | 1459   | 45.4        | 645    | 1  | DXS_RHIME   |
| 32         | 1455   | 45.2        | 629    | 1  | DXS_BACHD   |
| 33         | 1442   | 44.8        | 640    | 1  | DXS_SYMBL   |

## ALIGNMENTS

| RESULT 1 | ID   | DXS_VIBVU | STANDARD: | PRT: | 621 AA. |
|----------|--|-----------|-----------|------|---------|
| AC       | Q8DFR3   |           |           |      |         |
| DT       | 15-SEP-2003 (Rel. 42, Created)   |           |           |      |         |
| DT       | 15-SEP-2003 (Rel. 42, Last sequence update)  |           |           |      |         |
| DT       | 15-SEP-2003 (Rel. 42, Last annotation update)  |           |           |      |         |
| DE       | 1-deoxy-D-xylose-5-phosphate synthase (EC 2.2.1.7) (1-   |           |           |      |         |
| DE       | deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).   |           |           |      |         |
| GN       | DXS OR VY10315.  |           |           |      |         |
| OS       | Vibrio vulnificus.   |           |           |      |         |
| OC       | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  |           |           |      |         |
| OC       | Vibrionaceae; Vibrio.  |           |           |      |         |
| OX       | NCBI TaxID=672;  |           |           |      |         |
| RP       | SEQUENCE FROM N.A.   |           |           |      |         |
| RC       | STRAIN-CMCP6;  |           |           |      |         |
| RA       | Rhee U.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  |           |           |      |         |
| RA       | Choy H.E.;   |           |           |      |         |
| RT       | *Complete genome sequence of Vibrio vulnificus CMCP6.*;  |           |           |      |         |
| RL       | Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  |           |           |      |         |
| CC       | - FUNCTION: Catalyzes the acylol condensation reaction between C   |           |           |      |         |
| CC       | atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield  |           |           |      |         |
| CC       | 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).  |           |           |      |         |
| CC       | - CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-   |           |           |      |         |
| CC       | deoxy-D-xylose 5-phosphate + CO(2).  |           |           |      |         |
| CC       | - COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By   |           |           |      |         |
| CC       | similarity).   |           |           |      |         |
| CC       | - PATHWAY: Nucleoside diphosphate biosynthesis pathway; first step.  |           |           |      |         |
| CC       | - SUBUNIT: Homodimer (By similarity).  |           |           |      |         |
| CC       | - SIMILARITY: Belongs to the transketolase family. DXPS subfamily.   |           |           |      |         |
| CC       | - This SWISS-PROT entry is copyright. It is produced through a collaboration   |           |           |      |         |
| CC       | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |           |           |      |         |
| CC       | the European Bioinformatics Institute. There are no restrictions on its  |           |           |      |         |
| CC       | use by non-profit institutions as long as its content is in no way   |           |           |      |         |
| CC       | modified and this statement is not removed. Usage by and for commercial  |           |           |      |         |
| CC       | entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a> |           |           |      |         |
| CC       | or send an email to <a href="mailto:license@sdb-sdb.ch">license@sdb-sdb.ch</a> ).  |           |           |      |         |
| DR       | EMBL: AB016798; AAC08845.1; -  |           |           |      |         |
| DR       | HMAP: MF_00315; -; 1.  |           |           |      |         |
| DR       | InterPro: IPR005477; Dxs.  |           |           |      |         |
| DR       | InterPro: IPR005476; Transketolase_C.  |           |           |      |         |
| DR       | InterPro: IPR005475; Transketolase_N.  |           |           |      |         |
| DR       | InterPro: IPR005474; Transketolase_CK.   |           |           |      |         |
| DR       | Pfam: PF02780; transketolase_C; 1.   |           |           |      |         |
| DR       | Pfam: PF02779; transket pyr; 1.  |           |           |      |         |
| DR       | TIGRfam: TIGR00204; dxe; 1.  |           |           |      |         |
| DR       | PROSITE: PS00801; TRANSKETOLASE_1; 1.  |           |           |      |         |
| DR       | PROSITE: PS00802; TRANSKETOLASE_2; 1.  |           |           |      |         |
| KW       | Transketase; Flavoprotein; Thiamine pyrophosphate;   |           |           |      |         |
| KW       | Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.   |           |           |      |         |
| SQ       | SEQUENCE 621 AA; 67959 MW; 69DABA8B22B8D4B CRC64;  |           |           |      |         |

Query Match 65.7%; Score 2114; DB 1; Length 621;  
 Best Local Similarity 63.1%; Pred. No. 2,9e-141;  
 Matches 392; Conservative 102; Mismatches 123; Indels 4; Gaps 2;

```

QY 3 LTLD---YPLKNHTPADIRALSKDQLQGLADVRGYTHVTSISGGHFAAGCTVELT 59
DB 1 KTLIDSKPTTLALAEFDRLPKETLPTLDELRTYLINSVSSGSHLASGLTVELT 60
QY 60 VALHYVNTPVQDLVWDVGHQAYPHKILTRKERMPTIRTLGGVAFAPADESEYDAFV 119
DB 61 VALHYVNTPVQDLVWDVGHQAYPHKILTRKERMPTIRTLGGVAFAPADESEYDAFV 120
QY 120 GHSSSTISALGMAIAQOLGEGDKMVAIIGDSITGGAAYANMHADVYANLVIIND 179
DB 121 GHSSSTISALGMAIAQOLGEGDKMVAIIGDSITGGAAYANMHADVYANLVIIND 180
QY 180 NMSISPPVGMANNVITKVLSSKFSYSSVRESKALAKMPSVWELARKTEBEVKMIVPG 239
DB 181 NMSISPPVGMANNVITKVLSSKFSYSSVRESKALAKMPSVWELARKTEBEVKMIVPG 240
QY 240 TLFEELEFNYPIDGHDVEMVSTLENKDLTGPFVLAHVTKKGGYAPAEKOPLAYHG 299
DB 241 TLFEELEFNYPIDGHDVEMVSTLENKDLTGPFVLAHVTKKGGYAPAEKOPLAYHG 300
QY 300 VPAPDPTKDFPKAPAPSPHTYEVFRWLCDMAODERLGITPAMEGSGLVFEFSQKF 359
DB 301 VPAPDPTKDFPKAPAPSPHTYEVFRWLCDMAODERLGITPAMEGSGLVFEFSQKF 359
QY 360 PNRYPDVAIAEQHVAVTLANGAQCGAKPVVAIYSTFLOGYDQILHVALONLMLFALD 419
DB 360 PNRYPDVAIAEQHVAVTLANGAQCGAKPVVAIYSTFLOGYDQILHVALONLMLFALD 419
QY 420 RAGLVGDPDPTAGAFDVSVCIPNMLIMADNEBCKOMLTGPOHGGSPATYPPAK 479
DB 420 RAGLVGDPDPTAGAFDVSVCIPNMLIMADNEBCKOMLTGPOHGGSPATYPPAK 479
QY 480 GGGAAIDPTLTALTEIGKAEVRRHKSRIALANGSWTPVAEAKOIGATVVMREFVKEPD 539
DB 480 GGGAAIDPTLTALTEIGKAEVRRHKSRIALANGSWTPVAEAKOIGATVVMREFVKEPD 539
QY 540 QALVLELATHTVPTVEENVLAGANGSAINFLQAKVLMVCHGIDPRVBEQSGSEE 599
DB 540 QALVLELATHTVPTVEENVLAGANGSAINFLQAKVLMVCHGIDPRVBEQSGSEE 599
QY 540 ELLLEQLAHEHVDVLTTEENALAGAGAGVIEFMQEKIMKPVNLGIDPKFIHQGTQEE 599
DB 540 ELLLEQLAHEHVDVLTTEENALAGAGAGVIEFMQEKIMKPVNLGIDPKFIHQGTQEE 599
QY 600 LLSLVGDSKGIATIEFCFA 620
DB 600 LLSLVGDSKGIATIEFCFA 620
QY 600 LHEBLGIDKGIQALRHLYLA 620
DB 600 LHEBLGIDKGIQALRHLYLA 620

```

## RESULT 2

DXS\_VIBCH STANDARD; PRT; 626 AA.  
 AC 09KTL3;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-  
 deoxyxylose-5-phosphate synthase) (DXPS).  
 GN DXS OR VC0889.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OC NCBI\_TaxID=666;  
 RN 11;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 Dodson R.J., Hatt D.H., Hickey B.K., Peterson J.D., Unanue L.A.,  
 Gill S.R., Nelson K.E., Read T.D., Tellein H., Richtadson D.,  
 Ermolaeva M.D., Vamachewen J., Bass S., Qin H., Dragoi I., Sellers P.,  
 McDonald L., Uterback T., Fleischmann R.D., Nieman W.C., White O.,

```

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  

RA Fraser C.M.;  

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  

RT cholerae.";  

RL Nature 406:477-483 (2000).  

CC -1- FUNCTION: Catalyzes the acyloln condensation reaction between C  

CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield  

CC 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).  

CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-  

CC deoxy-D-xylose-5-phosphate + CO(2).  

CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By  

CC similarity).  

CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first step.  

CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first step.  

CC -1- SUBUNIT: Homodimer (By similarity).  

CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.  

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  

CC the European Bioinformatics Institute. There are no restrictions on its  

CC use. By non-profit institutions as long as its content is in no way  

CC modified and this statement is not removed. Usage by and for commercial  

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  

CC or send an email to license@isb-sib.ch).  

CC EMBL; AB004173; AAP94051.1; -  

CC PIR; H82266; H82266.  

CC TIGR; VC0889; -  

CC HAMAP; MF_00315; -; 1.  

CC InterPro; IPR005477; Dxs.  

CC InterPro; IPR005476; Transketolase C.  

CC InterPro; IPR005475; Transketolase CR.  

CC InterPro; IPR005474; Transketolase N.  

CC Pfam; PF02779; transket_pyr; 1.  

CC Pfam; PF02780; transketolase_C; 1.  

CC TIGRFAMs; TIGR00204; dxs; 1.  

CC PROSITE; PS00801; TRANSKETOLASE_1; 1.  

CC PROSITE; PS00802; TRANSKETOLASE_2; 1.  

CC Transferrase; Flavoprotein; Thiamine pyrophosphate;  

CC Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.  

CC SEQUENCE 626 AA; 68346 MW; 5B3BF9B9E851C9 CRC64;  


```

Query Match 65.2%; Score 2097.5; DB 1; Length 626;  
 Best Local Similarity 62.5%; Pred. No. 4,2e-140;  
 Matches 391; Conservative 109; Mismatches 117; Indels 9; Gaps 3;

```

QY 3 LTLD---YPLKNHTPADIRALSKDQLQGLADVRGYTHVTSISGGHFAAGCTVELT 59
DB 1 KTLIDSKPTTLALAEFDRLPKETLPTLDELRTYLINSVSSGSHLASGLTVELT 60
QY 60 VALHYVNTPVQDLVWDVGHQAYPHKILTRKERMPTIRTLGGVAFAPADESEYDAFV 119
DB 61 VALHYVNTPVQDLVWDVGHQAYPHKILTRKERMPTIRTLGGVAFAPADESEYDAFV 120
QY 120 GHSSSTISALGMAIAQOLGEGDKMVAIIGDSITGGAAYANMHADVYANLVIIND 179
DB 121 GHSSSTISALGMAIAQOLGEGDKMVAIIGDSITGGAAYANMHADVYANLVIIND 180
QY 180 NMSISPPVGMANNVITKVLSSKFSYSSVRESKALAKMPSVWELARKTEBEVKMIVPG 239
DB 181 NMSISPPVGMANNVITKVLSSKFSYSSVRESKALAKMPSVWELARKTEBEVKMIVPG 240
QY 240 TLFEELEFNYPIDGHDVEMVSTLENKDLTGPFVLAHVTKKGGYAPAEKOPLAYHG 299
DB 241 TLFEELEFNYPIDGHDVEMVSTLENKDLTGPFVLAHVTKKGGYAPAEKOPLAYHG 300
QY 300 VPAPDPTKDFPKAPAPSPHTYEVFRWLCDMAODERLGITPAMEGSGLVFEFSQKF 359
DB 301 VPAPDPTKDFPKAPAPSPHTYEVFRWLCDMAODERLGITPAMEGSGLVFEFSQKF 359
QY 360 PNRYPDVAIAEQHVAVTLANGAQCGAKPVVAIYSTFLOGYDQILHVALONLMLFALD 419
DB 360 PNRYPDVAIAEQHVAVTLANGAQCGAKPVVAIYSTFLOGYDQILHVALONLMLFALD 419

```

DB 360 PSQYDPAIAEONAVLATATGMAIAGHPVIAIYSTFLQGYDOLHDAVALMPLVWFAID 419  
 QY 420 RAGLVGPDPTAGADYSGYACIPNMLMAPADENEGOMLTGROHNGPASYVPRK 479  
 DB 420 RAGLVGADQOTGAGADLSYKICIPNMLMAPADENEGOMLTGROHNGPASYVPRK 479  
 QY 480 GCGAALDPTLTALEIGKAEVRH-----HGSRIAILAMGSMVTPAVEAGKOLGATVYNNR 534  
 DB 480 GCGVELESFPTALEIGKAEVRH-----HGSRIAILAMGSMVTPAVEAGKOLGATVYNNR 539  
 QY 535 VVPFOALVLELARTFVTVVEVTVVENVVAGAGSALITFLQAKVMPVONIGLPRFVQ 594  
 DB 540 VVPFDALVLELARTFVTVVEVTVVENVVAGAGSALITFLQAKVMPVONIGLPRFVQ 599  
 QY 595 GSRBELSLVGLDSKGLIATIQPCA 620  
 DB 600 GTOEHMELGIDGAGIERALIDYLA 625

RESULT 3  
 DXS\_PSEAE STANDARD; PRT; 627 AA.  
 ID DXS\_PSEAE STANDARD; PRT; 627 AA.  
 AC O9K6U7;  
 DT 16-OCT-2001 (Rel. 40, Last created)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DT 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-deoxyxylylose-5-phosphate synthase) (DXPS).  
 DB DXS OR PA044.  
 GN Pseudomonas aeruginosa.  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.  
 OC NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=20487113; PubMed=11034300;  
 RA Altincicek B., Hintz W., Sanderbrand S., Wiesner J., Beck E., Joma H.;  
 RT "Tools for discovery of inhibitors of the 1-deoxy-D-xylose 5-phosphate (DXP) synthase and DXP reductoisomerase: an approach with enzymes from the pathogenic bacterium Pseudomonas aeruginosa";  
 RL FEMS Microbiol. Lett. 190:329-333(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 15692 / PA01;  
 RA MEDLINE=20437377; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman F.S.L., Huinagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen";  
 RL Nature 406:959-964(2000).  
 RN [1]  
 RP FUNCTION: Catalyzes the acylol condensation reaction between C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylose-5-phosphate (DXP).  
 CC CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-deoxy-D-xylose 5-phosphate + CO(2).  
 CC COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By similarity).  
 CC PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.  
 CC PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first step.  
 CC SUBUNIT: Homodimer (By similarity).  
 CC SIMILARITY: Belongs to the transketolase family. DXPS subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, AF282878; AAF97240.1; --  
 CC EMBL, AE004821; AAG07431.1; --  
 CC DR EMBL; 683139; 683139.  
 CC DR HAMAP; MF\_00315; --; 1.  
 CC DR InterPro; IPR005477; Dxs.  
 CC DR InterPro; IPR005476; Transketolase\_C.  
 CC DR InterPro; IPR005475; Transketolase\_CR.  
 CC DR InterPro; IPR005474; Transketolase\_N.  
 CC DR Pfam; PF02779; transket\_pyr; 1.  
 CC DR Pfam; PF02780; transketolase\_C; 1.  
 CC DR TIGRfam; TIGR00204; dxs; 1.  
 CC DR PROSITE; PS00801; TRANSKETOLASE\_1; FALSE NEG.  
 CC DR PROSITE; PS00802; TRANSKETOLASE\_2; FALSE NEG.  
 CC DR Transferrase; Flavoprotein; Thiamine pyrophosphate;  
 CC Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.  
 CC KW  
 CC SEQUENCE 627 AA; 68049 MW; BBRP04296FFED57D CXC64;

Query Match 64.4%; Score 2069.5; DB 1; Length 627;  
 Best Local Similarity 63.4%; Pred. No. 3.9e-138;  
 Matches 388; Conservative 100; Mismatches 117; Indels 7; Gaps 2;  
 QY 8 PLKNHHPAPIRLALSKDQLQGLADEVRGYLTHVTSISGGHPAGIGTVELTVALHYFN 67  
 DB 16 FLDRASSPAELRRRGEADLTLDLQVLLVYVGTGSHFGAGLVVALLALHYFD 75  
 QY 68 TPVDQVWDVGHQAVPHKILTGKREKMPITRTIGVSAFPADESEYDAFVGHSSTIS 127  
 DB 76 TPDRVLVWDVGHQAVPHKILTEREELMGTLLQKQGLAFAFPADESEYDTPFGHSSTIS 135  
 QY 128 AALGMAIASQLRGEDEKRVAILIGGSGITGMAIYEMHAGVNNALVTLINDNKSIPP 187  
 DB 136 AALGMAIARLQGERKSVAVIGGALTAGMAFPAIINASVVDADMTLLINDNKSISEN 195  
 QY 188 VGAMNNVITKVLSSKPYSSVSEESKALAKMPSVWEILARKTEHYKGMIVGTLFEBIGF 247  
 DB 196 VGGLSNLYAKLISRTTISNMEBSKVLSTRPGAMEIARREHYAKGLVGTLPFBIGW 255  
 QY 248 NYFGEIDGHVEMVLTLENLKDVTGVPVHLVTVTKGKVAPAEKDPVLYGVPAFDPYK 307  
 DB 256 NYIGPIDGHDLPTVATLIRNRDMKGPQFELHVYTKGKGFAPALDPIGYHAI-----TK 310  
 QY 308 DFLPKAAP--SPHPTTEVPGRMICDMAADDERLIGTIPARESGGLVEFSQKPNRYFD 365  
 DB 311 LEAPGSAKKTGCPRTYSVFGOMCDMAQDARLIGITPAKESGDVAFSERYERYFD 370  
 QY 366 VALAEGAAVTLAAGACGAGAPVVAIYSTFLQGYDOLHDAVALONTDMFALDRAGLVG 425  
 DB 371 VALAEGAAVTLAAGACGAGAPVVAIYSTFLQRAYDOLHDAVALVQHLDFALDRAGLVG 430  
 QY 426 PDGPHAGAPFYSSVMEICIPNMLMAPADENEGOMLTGPOHGPASVRYRGKPGAAI 485  
 DB 431 EDGPHAGSPFISTURCIPGMLVMPSPDDELKRLITTYGLPDPBAVRYRGSGPNHPI 490  
 QY 486 DPTLTALEIGKAEVRHSGRIAILAMGSMVTPAVEAGKOLGATVYNNRFPDQALVE 545  
 DB 491 DPDLQPAVIGGVVRRRGVAVLVFGVLAEMAKVASELDATVVDKRFVPLDEALVRE 550  
 QY 546 LARTFVTVVEVTVVENVVAGAGSALITFLQAKVMPVONIGLPRFVQSRBELSLV 605  
 DB 551 LAGSHELLVLEENAVMVGAGSAGVGEPLASGELVPLDQGLPDIYVHAAPSEMLACG 610  
 QY 606 LDSKGLIATIEQ 617  
 DB 611 LDAAGIERAVQ 622

RESULT 4  
 DXS\_SHEON STANDARD; PRT; 622 AA.  
 ID DXS\_SHEON STANDARD; PRT; 622 AA.  
 AC Q8BGR9;

DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-  
 deoxyxylose-5-phosphate synthase) (DXP synthase) (DXPS).  
 GN DKS OR S01525.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NR-1.  
 RX MEDLINE=22297686; PubMed=12368913;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mehta B., Clayton R.A.,  
 RA Meyer T., Tappin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson R.J., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,  
 RA Feldguy T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123 (2002).  
 CC - FUNCTION: Catalyzes the acylotin condensation reaction between C  
 CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield  
 CC 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).  
 CC - CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-  
 CC deoxy-D-xylose-5-phosphate + CO(2).  
 CC - COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By  
 CC similarity).  
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.  
 CC - PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first  
 CC step.  
 CC - SUBUNIT: Homodimer (By similarity).  
 CC - SIMILARITY: Belongs to the transketolase family. DXPS subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AEO15598; AAN54586.1; -  
 CC TIGR: S01525; -  
 DR HAMAP: MF\_00315; - 1.  
 DR InterPro: IPR005477; Dxs.  
 DR InterPro: IPR005476; Transketolase\_C.  
 DR InterPro: IPR005475; Transketolase\_CR.  
 DR InterPro: IPR005474; Transketolase\_N.  
 DR Pfam: PF02780; transketolase\_C; 1.  
 DR Pfam: PF02779; transket\_pyr\_1.  
 DR TIGRfam: TIGR00264; dxs; 1.  
 DR PROSITE: PS00801; TRANSKETOLASE\_1; 1.  
 DR PROSITE: PS00802; TRANSKETOLASE\_2; 1.  
 DR Transferrase; Flavoprotein; Thiamine pyrophosphate;  
 KM Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.  
 SQ SEQUENCE 622 AA; 68083 MW; BE661C01BEA5EFB4 CRG4;  
 Query Match 63.5%; Score 2042; DB 1; Length 622;  
 Best Local Similarity 62.0%; Pred. No. 3,4e-136;  
 Matches 383; Conservative 105; Mismatches 126; Indels 4; Gaps 3;  
 QY 5 TDYPLKNTFPADIALSKDQQLADEVRYGLTFVTSISGSHFAGLTVETLTAALY 64  
 DB 6 SCPPVLAQANTPRLAKQLPOLALLPOLADERLRLSVGSSSHFASGIVELTALHY 65  
 QY 65 VFTPTVDQLVWDVGVQAVPKKILTKRKEMPTIRTKGVSAPAPADESEYDAFGVSHST 124  
 DB 66 VVNTPEDRLIWDVGVQAVPKKILTKRKEMPTIRTKGVSAPAPADESEYDAFGVSHST 125

QY 125 SISALAGNAISQLEGEEDKXWVAIIIGDSITGMAVYEAEMNAGDVANALVINDNMSI 184  
 DB 126 SISALMAVAABEKQARXVAVIIGDAMTGMGVFEAMTAGDILHNDMLVINDNMSI 185  
 QY 185 SPPVQANNTYLTKVISKFFSSVYRESKALAKPSVWELARKTEHKGNIIVETLLEE 244  
 DB 186 SENVVALNNHNLQAOLMSGFFYTIIRSSKRVKQGVFKEMAKRTEELKGVWVPEGLFEE 245  
 QY 245 IGFNFYCPIDGHDVEMLVSTLENLKDITGPVFLHVTYKKGKGYAPAEKDPALYHGPAPD 304  
 DB 246 IGFNFYCPIDGHDVEMLVSTLENLKDITGPVFLHVTYKKGKGYAPAEKDPALYHGPAPD 305  
 QY 305 PTXDPFLPAAPSPH-PTTTEVFGKMLCMAAODERLGITPAMRGSLVBSQCFPNRY 363  
 DB 306 PS--LEKPKATPKPGI-PTTSQVFGKMLCDIAQODEKVLGITPAMRGSGMVEFSORFPKQY 363  
 QY 364 FDVALAEQAVTLAAGQACQAKPVVAIYSTFLQRYQQLIHVALQMLDMFLALDRAGL 423  
 DB 364 FDVALAEQAVTLAAGQACQAKPVVAIYSTFLQRYQQLIHVALQMLDMFLALDRAGL 423  
 QY 424 VGPDPFTHAGAFDVSVMCIENMLIMADENECROWLTTFGPQH-GPASVRYPRKGP 482  
 DB 424 VQADGPTHQAFDLSFMRGCIENMVMAPSDEKCRQMLYTCYDAGPSAVRYPRGSAATG 483  
 QY 483 AAIIDPTLALALIGKAEVHHGSRITAILAWGSMVTPAVAGQOLGATVNMFPVXEPQAL 542  
 DB 484 ANQVAMTALPLFGKVIYRLGRILNMLNFTLNAALTAASLNDVVDVDFVXEPQAL 543  
 QY 543 VLELARTHDVTVYAEENYIAGAGSANTFLQAGKVLMPVGNIGLPSFVQGSRELLS 602  
 DB 544 VVEMAGTSDVTVYAEENYIAGAGSANTFLQAGKVLMPVGNIGLPSFVQGSRELLS 603  
 QY 603 LVGDSKGLIATIEPFA 620  
 DB 604 DLQIDAEKMLQINAYLA 621  
 RESULT 5  
 ID DKS\_YERPE STANDARD; PRT; 619 AA.  
 AC 08ZC45;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-  
 DE deoxyxylose-5-phosphate synthase) (DXP synthase) (DXPS).  
 GN DKS OR YP03177 OR Y1008.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parthill U., Wren B.W., Thomson N.R., Tildall R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltham T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moulton S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
 RL Nature 413:523-527 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIMS / Biovar Mediaevalis;  
 RX MEDLINE=22137863; PubMed=12142430;  
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,



entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AL627266; CAD08878.1; -

DR EMBL: AB016842; AAO70031.1; -

DR HAMAP: MF\_00315; -; 1.

DR InterPro: IPR005477; Dxs.

DR InterPro: IPR005476; Transketolase C.

DR InterPro: IPR005475; Transketolase CR.

DR InterPro: IPR005474; Transketolase\_N.

DR Pfam: PF02779; transket\_pyr; 1.

DR Pfam: PF02780; transketolase\_C; 1.

DR TIGRPFAM: TIGR00204; dxe; 1.

DR PROSITE: PS00801; TRANSKETOLASE\_1; 1.

DR PROSITE: PS00802; TRANSKETOLASE\_2; 1.

KW transferase; flavoprotein; thiamine pyrophosphate;  
Isoprene biosynthesis; thiamine biosynthesis; Complete proteome.

FT INIT MET 0 0 BY SIMILARITY

SQ SEQUENCE 619 AA; 67364 MW; E9DYC61243C29DB7 CRC64;

Query Match 62.6%; Score 2014.5; DB 1; Length 619;  
Best local Similarity 61.0%; Pred. No. 2,9e-134;  
Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;

QY 7 YPLKNIHTPADIRALSQDQLADVEVGYLTHTVVISGHPAAGLGVETLVALHYF 66  
DB 7 YPTLALVDSQTEIRLLPKESLPLCDLERRYLDSVRSRSGHFAAGLGVETLVALHYF 66

QY 67 NTFVDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFADSEYAFVGHSGSTSI 126  
DB 67 NTFPDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFADSEYAFVGHSGSTSI 126

QY 127 SAALGNALASQLRGEDEKXVAVITGDSITGMAVYAMNAGDVANLVLINDKMSISIP 186  
DB 127 SAGIGIAVAABKGGKXORRTWCVIGDGAITAGMAFEMNHGDIRPDMVLINDNEMSI8 186

QY 187 PVGAMNNYLTKLSSKRYSVRSKSKALAKMPSVWEIARKTEHVKMIVPGLFEEIG 246  
DB 187 NVGALNNHLLARLISGKLYSLRGGKVFSGVPITELKRIEHEIKKXVVPGLFEEIG 246

QY 247 FNFEGPIDGHVEMLVSTLENLKDGLGPELVAVYTKKKGVAERKOPLAHGVAPDPT 306  
DB 247 FNVIGPDGHVEMLVSTLENLKDGLGPELVAVYTKKKGVAERKOPITFHAVPKFDS 306

QY 307 KDELPKAAEPHPHYTVFGRMLCKMAOBERLLGTPARBSGLVFSOKENNYFVY 366  
DB 307 SGGLPKSS-GGLPKSSKIPDMICETAKOSKLMALTPARESGSWERSKTFPDYFV 365

QY 367 AIAEOHAVTLAAGACGAKFVVAIYSTPLOGGYDQLIHVALONTDMLFALDRAGIYGP 426  
DB 366 AIAEOHAVTLAAGIAGIYFVAIYSTPLOGGYDQLIHVALONTDMLFALDRAGIYGA 425

QY 427 DGPTHAGAFDYSTWECIPNMLMAPADENECROWLTGTFQDH-GPASVTRPRKGPRAI 485  
DB 426 DGOJTHOAFDYSTWECIPNMLMAPADENECROWLTGTFQDH-GPASVTRPRKGPRAI 485

QY 486 DPTTALIEIGAVRHHGSRILAIAMGSMYTPAVEAKGQDAGVYNNRFPYFQDQALVLE 545  
DB 486 TP-LEKPIKIGLYKRGEEKALILPGLIMPEAKVABALNATLYDMRFYFDDDTILE 544

QY 546 LARTHDVFTVEENVVIAAGAGSAINTPLQOKTLPVCIINGLPFRPEQSGRELLISVG 605  
DB 545 MAAGHDLVLILEENALIMGAGSGVNEVLMARKEVPVLTINGLPDFTIPQIGQEARLEIG 604

QY 606 LDSKGLATITROPCA 620  
DB 605 LDPAGIEKAKIAML 619

RESULT 7  
Dxs\_SALTY  
AC Q88RD1; STANDARD; PRT; 619 AA.

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DB 15-deoxy-D-xylylucose 5-phosphate synthase (EC 2.2.1.7) (1-deoxyxylylucose-5-phosphate synthase) (DXP synthase) (DXPs).

GN DKS OR STM0422.

OS Salmonella typhimurium.

CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

CC Enterobacteriaceae; Salmonella.

CC NCB1\_Taxid=602;

RP SEQUENCE FROM N.A.

RC STRAIN=L7 / SGC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lacleille P., Courtney L., Portolillo S., Ali J., Dante M., Du P., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";

RL Nature 413:852-856 (2001).

CC - FUNCTION: Catalyzes the acylotin condensation reaction between C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylylucose-5-phosphate (DXP) (By similarity).

CC - CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-deoxy-D-xylylucose 5-phosphate + CO(2).

CC - COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By similarity).

CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.

CC - PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first step.

CC - SUBUNIT: Homodimer (By similarity).

CC - SIMILARITY: Belongs to the transketolase family. DXPs subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE008715; AAL19376.1; -

DR Stycene; SG27277; dxe.

DR HAMAP: MF\_00315; -; 1.

DR InterPro: IPR005477; Dxs.

DR InterPro: IPR005476; Transketolase C.

DR InterPro: IPR005475; Transketolase CR.

DR InterPro: IPR005474; Transketolase\_N.

DR Pfam: PF02779; transket\_pyr; 1.

DR Pfam: PF02780; transketolase\_C; 1.

DR TIGRPFAM: TIGR00204; dxe; 1.

DR PROSITE: PS00801; TRANSKETOLASE\_1; 1.

DR PROSITE: PS00802; TRANSKETOLASE\_2; 1.

KW transferase; flavoprotein; thiamine pyrophosphate;  
Isoprene biosynthesis; thiamine biosynthesis; Complete proteome.

FT INIT MET 0 0 BY SIMILARITY

SQ SEQUENCE 619 AA; 67336 MW; FB3525B91DB6F346 CRC64;

Query Match 62.6%; Score 2013.5; DB 1; Length 619;  
Best local Similarity 61.0%; Pred. No. 3,4e-134;  
Matches 375; Conservative 106; Mismatches 131; Indels 3; Gaps 3;

QY 7 YPLKNIHTPADIRALSQDQLADVEVGYLTHTVVISGHPAAGLGVETLVALHYF 66  
DB 7 YPTLALVDSQTEIRLLPKESLPLCDLERRYLDSVRSRSGHFAAGLGVETLVALHYF 66

QY 67 NTFVDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFADSEYAFVGHSGSTSI 126  
DB 67 NTFPDQLVWDVGHQAPPHKILTGKREKMPITRTLGVSAPFADSEYAFVGHSGSTSI 126

QY 127 SAALGNALASQLRGEDEKXVAVITGDSITGMAVYAMNAGDVANLVLINDKMSISIP 186



```

DB 127 SAGIGIAVAABKEGDRRTVCITGCGAITAGAFEMAHAGDIRDMLVILINDMSISB 186
QY 187 PYGANNNTLVLSKFTSSVAREESKALAKRPSWELAKTHEEVKMINVGTLEELG 246
DB 187 NVGALNNHIAQLLSKLYSSLREGKRVFSGVPIKEILKRTHEEHIKAVVPGTLEELG 246
QY 247 FUYFPIIDHDVEMLVSTLENKDLGVPYLVHVTIKKGKGYAPAKEDPLAYHGPAPDPT 306
DB 247 FVYIGPVGDHVMGLISTLKNNRDLKGPFLIMTKKGGYEPKADPTITPAVYKPPDS 306
QY 307 KDFLPKAPSPHPTTEVFGRMLCDMAADERLLGITPANREGSGLVESQCFENKRYDY 366
DB 307 SCGLPKRSS-GGLPGYSKIFGDMICETAAADSKLMAITPMBGSGMVEFSRFTPPRYDY 365
QY 367 ATAEGHATVLAAGCQAGKAPVAIYSTLOSQYOLIHVALONLMDLFAIDRAGLVGP 426
DB 366 ATAEGHATVLAAGLIGKVPVVAIYSTLOSQYOLIHVALONLMDLFAIDRAGLVGP 425
QY 427 DGPTHAGAFDYSYMRCPNNMLTMAPADENECQMLTTFQHH-GRPASVYRPGKGPAAI 485
DB 426 DQTHOGAFDLSYLRCPIMVINTPSDENECQMLFTGYHVDGPTAVAYPRGNAQVYL 485
QY 486 DDTLTALEIGKAEVHRHSRIILAMGSWTFAVEAGKQIGATVNMKRPVPGDALVLE 545
DB 486 TP-LEKLPKIGKLVKKGSKLALINFTLMPBAKVAEALNLTVDKRFKLDITLIL 544
QY 546 LARTHDVFTVEENVLAGAGSAINFLQAKVLMPCVNIGLPDRFVBSBELLSLVG 605
DB 545 MAAGHDAVLTLENNIMGAGSGVNEVMAHRRKPVVNLITGLPDRFPGQEBARAEIG 604
QY 606 LDKGILATIEQCA 620
DB 605 LDAAGIEAKIKAMEA 619

RESULT 8
DVS_RALSO STANDARD; PRT; 636 AA.
ID DVS_RALSO STANDARD; PRT; 636 AA.
AC Q8XX95;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 1-deoxy-D-xylose 5-phosphate synthase (BC 2.2.1.7) (1-
deoxyxylose-5-phosphate synthase) (DXS synthase) (DXS).
DVS OR KSC221 OR RS01378.
GN Ralstonia solanacearum (pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_Taxid=305;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=EM11000;
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange R.,
RA Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigler P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
CC -1- FUNCTION: Catalyzes the acyloln condensation reaction between C
atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).
CC -1- CATALYTIC ACTIVITY: pyruvate + D-glyceraldehyde 3-phosphate = 1-
deoxy-D-xylose 5-phosphate + CO(2).
CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
step.
CC -1- SUBUNIT: Homodimer (By similarity).

```

```

CC -1- SIMILARITY: Belongs to the transketolase family. DXS subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: AL646068; CA015928.1; -.
DR HAMAP, MF_00315; -.
DR InterPro; IPR001017; Dehydrogenase_E1.
DR InterPro; IPR005477; Dxs.
DR InterPro; IPR005476; Transketolase_C.
DR InterPro; IPR005475; Transketolase_CR.
DR InterPro; IPR005474; Transketolase_N.
DR Pfam; PF00676; E1_dehydrog; 1.
DR Pfam; PF02729; transket_pyr; 1.
DR Pfam; PF02780; transketolase_C; 1.
DR TIGRfam; TIGR00204; dxs; 1.
DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
KW Transferrase; Flavoprotein; Thiamine pyrophosphate;
KW Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
SQ SEQUENCE 636 AA; 68300 MW; 819D1F42B9F1440B CMC64;

Query Match 62.5%; Score 2008.5; DB 1; Length 636;
Best Local Similarity 61.8%; Pred. No. 7.9e-14;
Matches 384; Conservative 92; Mismatches 134; Indels 11; Gaps 5;

QY 7 YELKNHTPADIALSKDQLADEVRGYLTIVTSISGHPAAGLTVELTVAHYVF 66
DB 3 YELATTIDAPAEILRLDRKQGLTADLRAPVLSVAOTGHLSSNLTVELTVAHYVF 62
QY 67 NIPVDLVWDVGHQAYPEKILITGRKERPTIRTLGVSAPFARDESRYDAFGVGHSSSTSI 126
DB 63 NTPDRIWVDVGHQGYPEKILITGRREGVATLRDLGSI6FRRSRSPDTEGTATSSSTSI 122
QY 127 SAALGMALASQLRGDKKMAIIGDSTIGMAYEAMNAGDV-NANLTVILINDMSIS 185
DB 123 SAALGMALGACTQGNRAIVAVIGDMSAGMAEAMNAGVTRNLPLVYVILINDMSIS 182
QY 186 PYGANNNTLVLSKFTSSVAREESKALAKRPSWELAKTHEEVKMINVGTLEELG 245
DB 183 PYGANNNTLVLSKFTSSVAREESKALAKRPSWELAKTHEEVKMINVGTLEELG 242
QY 246 GNNYGPIDGHDVEMLVSTLENKDLGVPYLVHVTIKKGKGYAPAKEDPLAYHGP 301
DB 243 GNNYGPIDGHDVEMLVSTLENKDLGVPYLVHVTIKKGKGYAPAKEDPLAYHGP 302
QY 302 APDPKDFLPKAPSPHPTTEVFGRMLCDMAADERLLGITPANREGSGLVESQCFEN 361
DB 303 KENPDEGKIPAPARA-KVSTYQVFGQMLCDMAADKRLVGLTPARREGSGMVEFSRFT 361
QY 362 RYFDVAIAEGHATVLAAGCQAGKAPVAIYSTLOSQYOLIHVALONLMDLFAIDR 421
DB 362 RYFDVAIAEGHATVLAAGCQAGKAPVAIYSTLOSQYOLIHVALONLMDLFAIDR 421
QY 422 GLVGPDPGTHAGAFDYSYMRCPNNMLTMAPADENECQMLTTFQHHGPASVYRPGK 481
DB 422 GLVGPDPGTHAGAFDYSYMRCPNNMLTMAPADENECQMLTTFQHHGPASVYRPGK 481
QY 482 GAIDPTTLTALEIGKAEVHRH-----HGRRIILAMGSWTFAVEAGKQIGATVNM 536
DB 482 GAIDPTTLTALEIGKAEVHRH-----HGRRIILAMGSWTFAVEAGKQIGATVNM 536
QY 537 PFDQALVLEIARTHDVFTVEENVLAGAGSAINFLQAKVLMPCVNIGLPDRFVBS 596
DB 542 PIDAAGVLEIARTHDVFTVEENVLAGAGSAINFLQAKVLMPCVNIGLPDRFVBS 601
QY 597 REELSLVGLDLSKGLATIEQ 617

```



DB 602 HPAALACGLDANGILASIRE 622

RESULT 9

ID DKS\_ECOLI STANDARD; PRT; 619 AA.

AC P77488; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 1-deoxy-D-xyulose 5-phosphate synthase (EC 2.2.1.7) (1-deoxyxyulose-5-phosphate synthase) (DXP synthase) (DXPS).

CN DKS OR B0420.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RA (1)

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC STRAIN=K12;

RX MEDLINE=98151473; PubMed=9482846;

RA Lohs L.M., Campos N., Rosa Putra S., Danielsen K., Rohmer M., Boronat A.,

RT "Cloning and characterization of a gene from Escherichia coli encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxyulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis."

RL Proc. Natl. Acad. Sci. U.S.A. 95:2105-2110(1998).

RM (2)

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis K.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.,

RT "The complete genome sequence of Escherichia coli K-12."

RL Science 277:1453-1474(1997).

RM (3)

RP SEQUENCE FROM N.A.

RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K., Duncan M., Federicci N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D., Nemeth A., Oefner P., Schramm S., Davis R.W.; Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.

RM (4)

RP SEQUENCE OF 1-6, AND CHARACTERIZATION.

RX MEDLINE=98058734; PubMed=9371765;

RA Sprenger G.A., Schorken U., Wiesger T., Grolle S., de Graaf A.A., Taylor S.V., Begley T.P., Bringer-Meyer S., Sahm H.;

RT "Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xyulose 5-phosphate precursor for isoprenoids, thiamin, and pyridoxol."

RL Proc. Natl. Acad. Sci. U.S.A. 94:12857-12862(1997).

RM (5)

RP CHARACTERIZATION.

RX MEDLINE=20115529; PubMed=10648511;

RA Kuzuyama T., Takagi M., Takahashi S., Seto H.;

RT "Cloning and characterization of 1-deoxy-D-xyulose 5-phosphate synthase from Streptomyces sp. strain CH190, which uses both the mevalonate and nonmevalonate pathways for isopentenyl diphosphate biosynthesis."

RL J. Bacteriol. 182:891-897(2000).

RT FUNCTION: Catalyzes the acylol condensation reaction between C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xyulose-5-phosphate (DXP).

CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-deoxy-D-xyulose 5-phosphate + CO(2).

CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By similarity).

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.

CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first step.

CC -1- SUBUNIT: Homodimer.

CC -1- MISCELLANEOUS: Optimal temperature is 42-44 degrees Celsius and optimal pH is 7.5-8.0.

CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF035440; AAC4162.1; -

DR EMBL; AE000148; AAC73523.1; -

DR EMBL; U82664; AAB40176.1; -

DR PIR; D64771; D64771.

DR EcoGene; E031612; dks.

DR HMAP; MF\_003125; -1.

DR InterPro; IPR005477; Dks.

DR InterPro; IPR005476; Transketolase\_C.

DR InterPro; IPR005475; Transketolase\_CR.

DR InterPro; IPR005474; Transketolase\_N.

DR Pfam; PF02779; transket pyr; 1.

DR Pfam; PF02780; transketolase\_C; 1.

DR TIGRFAMs; TIGR00204; dks; 1.

DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.

DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.

DR Transferrase; Flavoprotein; Thiamine pyrophosphate;

KW Isoprene biosynthesis; Thiamine biosynthesis; Manganese; Magnesium;

KW Complete proteome.

FT INIT MET 0

SQ SEQUENCE 619 AA; 67485 MW; D1808BEEF9DC9D2B CRC64;

Query Match 62.2%; Score 2000.5; DB 1; Length 619;

Best Local Similarity 60.2%; Pred. No. 2; Be-133;

Matches 370; Conservative 114; Mismatches 128; Indels 3; Gaps 3;

QY 7 YELKNIHTPADIDFALSCKDQQLQDLAEVRGLTFTVISGSGFAAGLGVETLVYALHYVF 66

DB 7 YETLALVDSTOELRLRKESLPKICDELRLYLDVSRSRSGHFMASGGLVETLVYALHYV 66

QY 67 MTPVDOLVMDGHOAYPHKILITGRKEMPTIRLTGVLSAPPARESEYDARFGVHSSTSI 126

DB 67 NTPPDOLIMDVGHOAYPHKILITGRKDKITIRKQGLHPPWRGSESDVLSVHSSTSI 126

QY 127 SAALGMALASGLRGEDKKMVALIGDSITGSMAYEAYANHAGDVANLVILINDMNSISP 186

DB 127 SAGGIAVAPAEKSKNRKRYVCVIGDALTACAFANHAGDIRPMLVILINDMNSISE 186

QY 187 PVGAMNNYLTVLSSKTYSSVREESKCALAMPSEWELARTTEHYKGMIVFGTLFEELG 246

DB 187 NVGALNNHLAGLISGLKYSSLRGGKVFSGVPIKEILAKTEHIIKGMVVPGLFEELG 246

QY 247 FNYGPIIDGHVEMLVNLTENLKDILGVPVFLHVVTKGKGYPAPAKPPLAHHGYPADPT 306

DB 247 FNYGIPVDGHVNLGITLTKMRDLKGFQFLHVVTKGKGYPAPAKPPLTHAHPKDPDS 306

QY 307 KDFPKAPSPHPYTVTFVFGMLCDMAQDERLLGITPAMESSGLVFESEKFPNRYTDV 366

DB 307 SGCLPKSS-GGLPFSYKFGDWLCEFAKDKMLAIFPAMESSGVMFESEKFPDRYEDV 365

QY 367 ALAEHNAVTLAAGAACQAKRVVAIYSTFLORGDLIHVVALNLDLFLPLDRAGLVGP 426

DB 366 ALAEHNAVTPAAGALIGKRTVAIYSTFLORAYDVAIDVAILKPLPLAIDRAGIVGA 425

QY 427 DGPTTAGAFDYSYKCAIPNMLINAPADENECROMLTGFGHH-GRASVYRPGKPGAAI 485

DB 426 DGQTHQGAFLDYLKCIPEVATIMTPEDENECROMLYTGYHNDGSPASVARYPRGNAVVEL 485

QY 486 DPTTLALBIGAEVPHHRSRLAIALANGSMTPPAVAGSQGLCATVNNRFPVDFDQALVLE 545

DB 486 TP-LEKLPIGKGIYVRREKLIATLFTLMPAPKAVASLNAVILDMKFAVPLDEALITLE 544

QY 546 LARTDVFVTEENVYIAGAGSAINTFLOAQKVLNFCVNIIGLDFVEQGREHLLSLVG 605  
 DB 545 MAASHEALVTYENNAIMCGAGGCVNEVIMARKKPPVYINIGLPPFFITQGTQGEEMRAELG 604  
 QY 606 LDSKGIATIBOCCA 620  
 DB 605 LDAGCEAKIKAMLA 619

RESULT 10  
 DKS\_ECOL6 STANDARD; PRT; 619 AA.  
 AC 08FKB9;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 2.2.1.7) (1-  
 deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).  
 DKS OR C0531.  
 GN Escherichia coli O6.  
 OS Escherichia coli O6.  
 OC Enterobacteriaceae; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 NCBI\_TaxID=217992;  
 (1)  
 SEQUENCE FROM N.A.  
 RP STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 Rashe D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,  
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -1- FUNCTION: Catalyzes the acyloln condensation reaction between C  
 atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield  
 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).  
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-  
 deoxy-D-xylulose 5-phosphate + CO(2).  
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By  
 similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.  
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first  
 step.  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EXBL: AEO16756; AAN79009.1; -.  
 DR HAMAB: MF 00315; -; 1.  
 DR InterPro: IPR005477; Dxs.  
 DR InterPro: IPR005476; Transketolase\_C.  
 DR InterPro: IPR005475; Transketolase\_CR.  
 DR InterPro: IPR005474; Transketolase\_N.  
 DR Pfam: PF02780; transketolase\_C; 1.  
 DR Pfam: PF02779; transket\_Pyr; 1.  
 DR TIGRfam: TIGR00204; dks1.1.  
 DR PROSITE: PS00801; TRANSKETOLASE\_1; 1.  
 DR PROSITE: PS00802; TRANSKETOLASE\_2; 1.  
 RA Transferrase; Flavoprotein; Thiamine pyrophosphate;  
 RA isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.  
 KW INIT MET 0 BY SIMILARITY.  
 FT INIT MET 0 BY SIMILARITY.  
 SQ SEQUENCE 619 AA; 67501 MW; E2AADIAS64FCG30 CRC64;

Query Match 62.2%; Score 1999.5; DB 1; Length 619;  
 Best Local Similarity 60.0%; Pred. No. 3.3e-133;

Matches 369; Conservative 115; Mismatches 128; Indels 3; Gaps 3;  
 QY 7 YPLIKNIHPADIRALSKDOQLQDLADEVRGTLTHTVSISGHPAAGLGVETLVAHYV 66  
 DB 7 YPTLALVDSQBELLPKESLPKCDLRRYILDSVRSRSGHPASGATVELVTAALHVY 66  
 QY 67 NTPVDOLVMDVGEQAYHKKILTKGKEMPTIRLTGVSAPAPADESRDAGVCHSTSI 126  
 DB 67 NTPPDOLIMDVGEQAYHKKILTKGRDKIGTRQGLHPPFWRGESRDVLSVGHSTSI 126  
 QY 127 SAALGMAIASQLRSDKQVALLIGDSITGMAEYANNAHDVANNILVILNDNMSISP 186  
 DB 127 SAGIGIAVAPAKRKRRTYCVIGDAITAGMAEYANNAHDINPDLVYLNDEMISSE 186  
 QY 187 PVGANNVLTIVISSKFTSYVRESKATALAKSVEWELAKRTSEHYKMTYVPTLFEIS 246  
 DB 187 NVGALNNHLAKQLLSGKLYSSLRBEGKTVFSGVPIKEILKRTSEHIKGMVVPGLFBEIS 246  
 QY 247 FNYGEPIDGHEMLVSTLENTKDLTGVPFLHYVTKKGGYAPAKDPLAHGVAPFDP 306  
 DB 247 FNYIGPVDGHDVLGLITLKNMRDLKQFPLHMTKGRGTEPAEKDPIETHAIPKDP 306  
 QY 307 KDFPKAAPSPHYTYTFVGRMLCMAAODERLIGITPAMEGSGLVESQKTPNRYDV 366  
 DB 307 SGCLPKPS -GGLPSYSKLPEDWLCEYAKDKMAITPAMEGSGMVEPSKFPDRYDV 365  
 QY 367 ALAAGHAYTLAAGACGAKPVAYISTYTORGDOLIHVYALQNTDMLPALDRAGLV 426  
 DB 366 ALAAGHAYTLAAGALGKPVAYISTYTORGDOLIHVYALQNTDMLPALDRAGLV 425  
 QY 427 DGPTAGAFDYSYVRCIPNMLINAPADENECROMLITGPOHH-GRASVYPRGKPGAI 485  
 DB 426 DGQHQGAFDLSYRCIPENYIMTPSDNECRQVLYGHNDPSAVRYRGNVGL 485  
 QY 486 DPTTALTEIGFAYRHGSRITAIAGSMVTPPVEAGKQLGATVYNNRFPVPOATLE 545  
 DB 486 TP-LEKPIGIGIVRRREKKAIIIFGLMEPAKVAESINATLVDMKFEVPLDLS 544  
 QY 546 LARTDVFVTEENVYIAGAGSAINTFLOAQKVLNFCVNIIGLDFVEQGREHLLSLVG 605  
 DB 545 MAASHEALVTYENNAIMCGAGGCVNEVIMARKKPPVYINIGLPPFFITQGTQGEEMRAELG 604

RESULT 11  
 DKS\_ECOL57 STANDARD; PRT; 619 AA.  
 ID DKS\_ECOL57  
 AC 08X876;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 2.2.1.7) (1-  
 deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).  
 GN DKS OR 20523 OR EC50474.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 NCBI\_TaxID=83334;  
 (1)  
 SEQUENCE FROM N.A.  
 RP STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Mayhew G.F., Burland V., Mau B., Glasner J.D.,  
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 Grodzick E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,  
 Grodeck J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 Weich R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RA Nature 409:529-533(2001).

RN SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / RIMD 0509952;  
 RA MEDLINE=21156231; PubMed=11258796;  
 RX Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 Rida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
 Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 RA "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 CC -1- FUNCTION: Catalyzes the acylol condensation reaction between C  
 atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield  
 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).  
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-  
 deoxy-D-xylulose 5-phosphate + CO(2)  
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By  
 similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.  
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first  
 step.  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AE005221; AAG54770.1; -  
 DR EMBL; AP002551; BAB33897.1; -  
 DR PIR; B90688; B90688.  
 DR PIR; F85538; F85538.  
 DR HAMAP; MF\_00315; -; 1.  
 DR InterPro; IPR005477; Dxs.  
 DR InterPro; IPR005476; Transketolase\_C.  
 DR InterPro; IPR005475; Transketolase\_CR.  
 DR InterPro; IPR005474; Transketolase\_N.  
 DR Pfam; PF02779; transket\_pyr; 1.  
 DR Pfam; PF02780; transketolase\_C; 1.  
 DR TIGRfam; TIGR00204; dxs; 1.  
 DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.  
 DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.  
 DR Transferrase; Flavoprotein; Thiamine pyrophosphate;  
 KW Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.  
 FT INIT MET 0 BY SIMILARITY.  
 FT SEQUENCE 619 AA; 67501 MW; F8150610E6FFC20 CRC64;  
 SQ  
 Query Match 62.0%; Score 1995.5; DB 1; Length 619;  
 Best Local Similarity 59.8%; Pred. No. 6.3e-133;  
 Matches 368; Conservative 115; Mismatch 129; Indels 3; Gaps 3;

Db 247 FNYIGPVDGHDVGLITTLTKNNRDLKGPQFLHMTKXGGRYPAEKDPITTHAVPKFDP 306  
 Qy 307 KDFLPRAAPSPHPPTTEVEGRMCDMAQDERLGLITPAKBSGGLVSEFQKFPNRYFDV 366  
 Db 307 SGLCPKSS-GGLPSYKTRIDMCTRAADNKLMTTPAKBSGGLVSEFQKFPNRYFDV 365  
 Qy 367 AIAEONAVTLAAGACQAGAPVALYSTPLQGYDQIHDAVQNTDMLPALRAGIVCP 426  
 Db 366 AIAEONAVTLAAGACQAGAPVALYSTPLQGYDQIHDAVQNTDMLPALRAGIVCP 425  
 Qy 427 DGPHTGAEFYSYMRCPNMLTMAPADENECRMILTGPQSH-SPASVRYPRGKGPAAI 485  
 Db 426 DQTHQCAFSLTKRCLPEKMTPLTSPDENCRMILTGHYHNGPSAVRYPRGKGPAAI 485  
 Qy 486 DPTVTALEIKARVRHGRSHRIATLWGSVTPAVRAGKQIGATVNNRFPKPDQALVE 545  
 Db 486 TP-LEKPIKGIYKRGKGLITLFTLMPPEAKVAESINATLVMPRFPKPDLETLLE 544  
 Qy 546 LARHDVFTVVERVAVAGGSAINTFLOKXVLMPCNIGLDRFVQSGREBELSLVG 605  
 Db 545 MAASHEALVTEERVAIWGAGSGVNEVMAHRRKVPVNLIGLDPFLPQGTQSEMRALG 604  
 Qy 606 LDSKGLATTEOPCA 620  
 Db 605 LDMAGMEAKIKAMIA 619  
 RESULT 12  
 Dxs\_XANCP STANDARD; PRT; 638 AA.  
 AC 08P815;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xylulose 5-phosphate synthase (BC 2.2.1.7) (1-  
 deoxyxylulose-5-phosphate synthase) (Dxs synthase) (DXPS).  
 GN Dxs OR XCC2434.  
 OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=340;  
 (1)  
 SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33913 / NCPPB 528;  
 RX MEDLINE=22022145; PubMed=12024217;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
 Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,  
 Camarotte G., Candavan F., Cardoso J., Chamego F., Ciapina L.P.,  
 Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
 Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
 Forughieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 Katayama A.M., Kisli L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 Locelli E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 Martins E.C., Meidanis U., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 Pereira H.A., Rossi A., Sema J.A.D., Silva C., de Souza R.F.,  
 Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,  
 Setubal J.C., Kitajima J.P.;  
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing  
 host specificities.";  
 RL Nature 417:459-463(2002).  
 CC -1- FUNCTION: Catalyzes the acylol condensation reaction between C  
 atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield  
 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).  
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-  
 deoxy-D-xylulose 5-phosphate + CO(2).  
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By  
 similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.  
 CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first  
 step.

```

RESULT 13
DMS_XANAC
ID DMS_XANAC STANDARD; PRF: 638 AA.
AC Q8P8T7;
DT 28-FEB-2003 (rel. 41, Created)
DT 28-FEB-2003 (rel. 41, Last sequence update)
DT 15-SEP-2003 (rel. 42, Last annotation update)
DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-
DMS OR XAC2565.
XN Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OC NCBI_TaxId=92829;
[1]
RP SEQUENCE FROM N.A.
RC SRRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reimach F.C., Farah C.S., Purlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Cacciatore G., Camavau F., Cardozo J.F., Chabergo F., Cipina L.P.,
RA Ciccarelli R.M.B., Coutinho L.B., Curcio-Santos J.R., El-Dorri H.,
RA Peria J.B., Ferreira A.J.S., Ferreira R.C.C., Gerber A.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gubert A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Melians J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sema J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
RA Setubal J.C., Kitejima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
CC -1- FUNCTION: Catalyzes the acyloln condensation reaction between C
CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
CC 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).
CC -1- CATALYTIC ACTIVITY: pyruvate + D-glyceraldehyde 3-phosphate = 1-
CC deoxy-D-xylose 5-phosphate + CO(2).
CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By
CC similarity).
CC -1- PATHWAY: Biomembrane terpenoid biosynthesis pathway; first step.
CC -1- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
CC step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation in
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb.ch).
CC -----
CC EMBL, AB011897; AAM37415.1; -.
CC HAAAP; MF_00315; -. 1.
CC InterPro; IPR005477; Dxs.
CC InterPro; IPR005476; Transketolase_C.
CC InterPro; IPR005475; Transketolase_CR.
CC InterPro; IPR005474; Transketolase_N.
CC Pfam; PR02779; transket pyr. 1.
CC Pfam; PR02780; transketolase_C. 1.
CC TIGRFAMs; TIGR00204; dxe. 1.
CC DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
CC DR PROSITE; PS00802; TRANSKETOLASE_2; FALSE_NEG.
CC Transferrase; Flavoprotein; Thiamine pyrophosphate;
CC Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
CC SEQUENCE 638 AA; 66479 MW; 3D739EBE91C17CF7 CRC64;

```

Query Match 61.8%; Score 1966.5; DB 1; Length 638;  
 Best Local Similarity 61.4%; Pred. No. 2.8e-132;  
 Matches 378; Conservative 91; Mismatches 142; Indels 5; Gaps 3;

4 TTDYPLKNIHTPADIRALSKDQLOQLADEVRGTLTHYVSGHFAAGLGVLTVALH 63  
 4 TTYPRLSRIQTDPDDRPEAEHLTAIEELRSYLIESVSGGHFAAGLGVLTVALH 63  
 4 YVNTPTDQVWVNGVQVAPHKILITGKEMPTIRLGGISAPPADESYDAFGVGHSS 123  
 64 YLYQTPVDQVWVNGVQVAPHKILITGKEMPTIRLGGISAPPADESYDAFGVGHSS 123  
 124 TSISAAIGMAIASQREDEKCKVAIIGDSITGGMAYEAMNHG--DVANALLVILNDMD 181  
 124 TSISAAIGMAIAORNDDEKVAIVIGDMTGMVTEALNHGMDPEKMLVILNDNR 183  
 182 MSISPPVGAANNVITKYLSSKFTYSVEESKALA--KMSVWELAKTBEHYKAVIQ 239  
 184 MSISEAVGGLTKMLGRASGSRITNAIREGSKITLGDKNPITAFYRFRWEEHWKGFVPS 243  
 240 TLEHIGVNYFGPIGDGVEMLVSTLENKDLGVPFLVYVTKKGAYAPAECDPLAHS 299  
 244 TLEFEMGPHYGTPIGDGLPSLVGALKTLQTLGPPQLAHVITTKGKYLHABGQIGFHA 303  
 300 VPAPDPKDELPRKAPSPHYTEVEFGMLCKMAAODERLLGITPARESSGGLVSESQKE 359  
 304 VGPPDPKGLVAKTG-AKKPTIYDVFSDDWCMAAARPKLVITTPARBSGGLVRSKQV 362  
 360 PNRVFDVAIAEQAHTLAAGACOGAFVVAIISTIGRGYDQILHDVALQNDMLPADD 419  
 363 PORYFDVAIAEQAHTLAAGACOGAFVVAIISTIGRGYDQILHDVALQNDMLPADD 422  
 420 RAGLVGDPGPHAFAPYSYMRGIPNMLIAPADENRCROMLTGFGHHPASVAPRGRK 479  
 423 RGVGVGDPGATHAGNDLSFRCVFNHVAHPADBAECROMLTGFGYGPAAVRPRGR 482  
 480 GPGAAIDPTLTALTEIGRAEVAHHSRIALAMGSMTPPAVEAGKQGLATVYNNRPVYPPD 539  
 483 GPGAAIDPTLTALTEIGRAEVAHHSRIALAMGSMTPPAVEAGKQGLATVYNNRPVYPPD 542  
 540 QALVTEIARFADVAVTEENVVIAAGASAIINTPLQOKNMPVYNIGLPRFPGQSRRE 599  
 543 KALLEIATKATGDTVTEEDVNVAGAGSVSELINAEALITPMLHLPLDPSFQHHASRED 602  
 600 LSLVGLDSKGLIATI 615  
 603 LLAERAGIDQAGIRAL 618

RESULT 14  
 DLS\_PASMTU STANDARD; PRT; 614 AA.

AC P57848;  
 DT 16-OCT-2001 (Rel. 40, Last Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xylose 5-phosphate synthase (EC 2.2.1.7) (1-  
 deoxyxylose-5-phosphate synthase) (DXP synthase) (DXPS).  
 GN DKS OR PM0532.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 OK NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Pm70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 "Complete genomic sequence of Pasteurella multocida Pm70."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 CC -1- FUNCTION: Catalyzes the acyloin condensation reaction between C  
 atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield

CC 1-deoxy-D-xylose-5-phosphate (DXP) (By similarity).  
 CC -1- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-  
 CC deoxy-D-xylose 5-phosphate + CO(2).  
 CC -1- COFACTOR: Binds 1 thiamine pyrophosphate per subunit (By  
 CC similarity).  
 CC -1- PATHWAY: Nucleonate terpenoid biosynthesis pathway; first step.  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SIMILARITY: Belongs to the transketolase family. DXPS subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE006088; AK02616.1; -.  
 DR HAMAP; MF 00315; -; 1.  
 DR InterPro; IPR005477; Dxs.  
 DR InterPro; IPR005476; Transketolase\_C.  
 DR InterPro; IPR005475; Transketolase\_CR.  
 DR InterPro; IPR005474; Transketolase\_N.  
 DR Pfam; PF02779; transket\_pyr; 1.  
 DR Pfam; PF02780; transketolase\_C; 1.  
 DR TIGRPFAM; TIGR00204; dxs; 1.  
 DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.  
 DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.  
 DR Transferrase; Flavoprotein; Thiamine pyrophosphate;  
 KM Isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.  
 SQ Sequence 614 AA; 67426 MW; EBSYD95DBBCAD3 CR64;

Query Match 61.2%; Score 1969; DB 1; Length 614;  
 Best Local Similarity 60.1%; Pred. No. 4.5e-131;  
 Matches 369; Conservative 108; Mismatches 133; Indels 4; Gaps 3;

6 DYPLKNIHTPADIRALSKDQLOQLADEVRGTLTHYVSGHFAAGLGVLTVALH 65  
 3 NYPLSLINSPELDLISKEQDPOICERATLSSVSQSGHLSGLATVLTVALH 62  
 66 FNTPVQOLVWDVGHQAYPKKILITGRKEMPTIRLGGISAPPADESYDAFGVGHSS 125  
 63 YKTPDQVWVNGVQVAPHKILITGKEMPTIRLGGISAPPADESYDAFGVGHSS 122  
 126 ISALGMAIASQREDEKCKVAIIGDSITGGMAYEAMNHGADVANALLVILNDMSIS 185  
 123 ISAGLGAIAAQRNAGRTVCVIGDAITAGAFEMNHAGALHTDMLVILNDMSIS 182  
 186 PPIVGAANNVITKYLSSKFTYSVEESKALA--KMSVWELAKTBEHYKAVIQ-GTLFEE 244  
 183 ENVGALNNIARLLTGSFYSIRSGKILSGMPPIKEVKTIEHYKAVIQFVPGVIMFQ 242  
 245 LGFNYPGPIGDGVEMLVSTLENKDLGVPFLVYVTKKGAYAPAECDPLAHS 304  
 243 LGFNYPGPIGDGVEMLVSTLENKDLGVPFLVYVTKKGAYAPAECDPLAHS 302  
 305 PTKDPLKAPSPHYTEVEFGMLCKMAAODERLLGITPARESSGGLVSESQKE 364  
 304 PTKDPLKAPSPHYTEVEFGMLCKMAAODERLLGITPARESSGGLVSESQKE 360  
 303 HLSGOLPKS--NTTPYSKIPGDMWLCENAKENPKIGITPARESSGGLVSESQKE 360  
 365 DVAAIEQAHTVLAAGACOGAFVVAIISTIGRGYDQILHDVALQNDMLPADD 424  
 361 DVAAIEQAHTVLAAGACOGAFVVAIISTIGRGYDQILHDVALQNDMLPADD 420  
 425 GPDGPHAFAPYSYMRGIPNMLIAPADENRCROMLTGFGHHPASVAPRGRK 484  
 421 GADGQTHQAGFISFRCVFNHVAHPADBAECROMLTGFGYGPAAVRPRGR 480  
 485 IDPTLTALTEIGRAEVAHHSRIALAMGSMTPPAVEAGKQGLATVYNNRPVYPPD 544  
 481 LTP-LTMLAGSNLIRSEKIALINFTGLTPTALINVAEKAAVYIDMFVPIOVERIH 539





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 46.2142 Seconds  
(without alignments)  
3461.979 Million cell updates/sec

Title: US-09-941-947a-6

Perfect score: 3216  
Sequence: 1 MKLITDYPFLKNHTPADIR.....LSLVGLDSKGLIATIEQPCA 620

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_ricent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriag:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1          | 2114   | 65.7        | 621    | 16    | Q8DFR3      |
| 2          | 2042   | 63.5        | 622    | 16    | Q8EGR3      |
| 3          | 1999.5 | 62.2        | 620    | 16    | Q8FKB9      |
| 4          | 1795   | 55.8        | 626    | 16    | Q8D357      |
| 5          | 1592.5 | 49.5        | 643    | 16    | Q8G292      |
| 6          | 1498.5 | 46.6        | 644    | 16    | Q8KFI9      |
| 7          | 1482.5 | 46.1        | 716    | 10    | Q8D174      |
| 8          | 1445   | 45.2        | 655    | 16    | Q8D174      |
| 9          | 1444.5 | 44.3        | 711    | 10    | Q8L692      |
| 10         | 1440.5 | 44.8        | 715    | 10    | Q8L6C6      |
| 11         | 1437.5 | 44.7        | 630    | 2     | Q8K213      |
| 12         | 1435.5 | 44.6        | 709    | 10    | Q9PSG2      |
| 13         | 1435.5 | 44.6        | 713    | 10    | Q9PSG5      |
| 14         | 1433   | 44.6        | 636    | 2     | Q8GKAO      |
| 15         | 1430.5 | 44.5        | 722    | 2     | Q8W4V1      |
| 16         | 1429.5 | 44.4        | 717    | 10    | Q8L693      |

|    |        |      |      |    |        |
|----|--------|------|------|----|--------|
| 17 | 1418.5 | 44.1 | 724  | 10 | Q64904 |
| 18 | 1408.5 | 43.8 | 725  | 10 | Q9FV39 |
| 19 | 1402.5 | 43.6 | 719  | 10 | Q9XHS0 |
| 20 | 1402   | 43.5 | 735  | 10 | Q81954 |
| 21 | 1397.5 | 43.5 | 719  | 8  | Q78328 |
| 22 | 1311.5 | 40.8 | 604  | 10 | Q9LVP5 |
| 23 | 1278.5 | 39.8 | 628  | 10 | Q8LPS4 |
| 24 | 1235.5 | 38.4 | 634  | 16 | Q8P153 |
| 25 | 1131.5 | 35.2 | 616  | 16 | Q8P152 |
| 26 | 1111.5 | 34.6 | 649  | 2  | Q9F1V2 |
| 27 | 1097   | 34.1 | 613  | 2  | Q8VUR8 |
| 28 | 1078.5 | 33.5 | 642  | 16 | Q8CJP7 |
| 29 | 1062.5 | 33.0 | 628  | 10 | Q9SK01 |
| 30 | 1053   | 32.7 | 700  | 10 | Q9LEF9 |
| 31 | 1030   | 32.0 | 600  | 16 | Q8R639 |
| 32 | 892.5  | 27.8 | 586  | 16 | Q97TJ5 |
| 33 | 886    | 27.5 | 580  | 16 | Q9CF08 |
| 34 | 882.5  | 27.4 | 599  | 16 | Q9CFJ5 |
| 35 | 870    | 27.1 | 583  | 16 | Q8R606 |
| 36 | 854    | 26.6 | 354  | 2  | Q69774 |
| 37 | 852.5  | 26.5 | 536  | 16 | Q50408 |
| 38 | 848.5  | 26.4 | 428  | 10 | Q944G7 |
| 39 | 823    | 25.6 | 406  | 10 | Q94CE7 |
| 40 | 819.5  | 25.5 | 1205 | 5  | Q96694 |
| 41 | 819.5  | 25.5 | 1205 | 5  | Q8IDM0 |
| 42 | 803.5  | 25.0 | 612  | 16 | Q8EKK7 |
| 43 | 689    | 21.4 | 774  | 10 | Q8G585 |
| 44 | 615    | 19.1 | 294  | 10 | Q9FEV6 |
| 45 | 522    | 16.2 | 272  | 2  | Q9WXP0 |

## ALIGNMENTS

## RESULT 1

ID Q8DFR3 PRELIMINARY; PRT; 621 AA.  
AC Q8DFR3:  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Deoxyxylulose-5-phosphate synthase.  
GN VV10315.  
OS Vibrio vulnificus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrrio.  
OC NCBI\_taxid=672;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CMCP6;  
RA Rhoe J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.,  
RT "Complete genome sequence of *Vibrio vulnificus* CMCP6,"  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016798; AAC08845.1; -  
KW Complete proteome.  
SQ SEQUENCE 621 AA; 67959 MM; 69DABA8BH228BD4B CMC64;

|    |                       |   |                     |                 |             |
|----|-----------------------|---|---------------------|-----------------|-------------|
| QY | Query Match           | 65.7%   | Score 2114;         | DB 16;          | Length 621; |
| QY | Best Local Similarity | 63.1%   | Pred. No. 1.4e-145; |                 |             |
| QY | Matches               | 392;  | Conservative 102;   | Mismatches 123; | Indels 4;   |
| QY |                       |   |                     | Gaps            | 2;          |
| DB | 3                     | LTTP---YELKNTHTPADIRALSKDQLOQLADEVRGYLHTTVISIGGHPAGIGTVELT  | 59                  |                 |             |
| DB | 1                     | MTLIDSKRPTLALATETPOHRLPLRETLPTLCDRLRYLANSVSQSSGHLASGIGTVELT | 60                  |                 |             |
| QY | 60                    | VALHYVFTPTDQLVMPYVGHQAYPHKILTGKEMPTIRLIGVSAFAPRDESETDAGV    | 119                 |                 |             |
| DB | 61                    | VALHYVFTPTDQLVMPYVGHQAYPHKILTGKEMPTIRLIGVSAFAPRDESETDAGV    | 120                 |                 |             |
| QY | 120                   | GHSSTISALGMAIISQIRGEGDKMVAIIGDSITGMAVEAMNHAGDVANLVLIND      | 179                 |                 |             |
| DB | 121                   | GHSSTISALGMAIISQIRGEGDKMVAIIGDSITGMAVEAMNHAGDVANLVLIND      | 180                 |                 |             |



QY 180 NMGISPPVGMANNLTITKVSRYSSVRESKRLAKMPVWEIARTEEHKMGTVRG 239  
 DB 191 NEMTSSSEVGAANNHIAVLGNLTSTIREGKVLGSGVPPKELVRRTEHLLKGVNPG 240  
 QY 240 TLFEELGFRNYGPIIDHGVEMLVSTLENLKDLTGPFVHVYTKKGGYAPAEKDPAYHG 299  
 DB 241 TLFEELGFRNYGPIIDHGVEMLVSTLENLKDLTGPFVHVYTKKGGYAPAEKDPAYHG 300  
 QY 300 VPAFPTKDFLPKAPSPHPTTTEVFGKWLCDMAADDERLLGTPAMRESGGLVPSQKF 359  
 DB 301 VKRFPANHSILPKSS-GGKPSFSNIFGDFLCDMAADDERLLGTPAMRESGGLVPSQKF 359  
 QY 360 PRRYDVVAIAEQHATVLAAGACOGAKPVVAIYSTFLORGVDLIHDVALQNLMLFALD 419  
 DB 360 PRRYDVVAIAEQHATVLAAGACOGAKPVVAIYSTFLORGVDLIHDVALQNLMLFALD 419  
 QY 420 RAGLVGPDGPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 479  
 DB 420 RAGLVGPDGPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 479  
 QY 480 GGGTPESEFTTLEIGKGLVROGKRLVQSEKVALSFTGLTANLAEALNATVADMTVRLD 539  
 DB 480 GGGTPESEFTTLEIGKGLVROGKRLVQSEKVALSFTGLTANLAEALNATVADMTVRLD 539  
 QY 540 QALVLELATHDVFTVEENVVLAGAGSAINTELQAKVLMPCVNIQLPDRFVEGSEEE 599  
 DB 540 QALVLELATHDVFTVEENVVLAGAGSAINTELQAKVLMPCVNIQLPDRFVEGSEEE 599  
 QY 600 LSLVGLDSKGLATTEQFCA 620  
 DB 600 LSLVGLDSKGLATTEQFCA 620

RESULT 2  
 QY 08EGR9 PRELIMINARY; PRT; 622 AA.  
 DB 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 GN DKS OR S01525.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OC NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NR-1.  
 RX MEDLINE=2237686; PubMed=12368813;  
 RA Heideberg J.F., Paulsen I.T., Nelson K.E., Galois E.J., Nelson W.C.,  
 Read T.D., Bisen J.A., Seshadri R., Ward N., Mehe B., Clayton R.A.,  
 Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty J.F.,  
 DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 Vamathevan J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,  
 Mueller J., Kfoury H., Gill J., Uterback T.R., McDonald L.A.,  
 Feldlyum T.V., Smith H.O., Venter J.C., Neillson K.H., Fraser C.M.,  
 "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 Shewanella oneidensis."  
 RT Nat. Biotechnol. 20:1118-1123 (2002).  
 RL EMBL; AB015598; AAN54586.1; -.  
 DR TIGR; S01525; -.  
 KW Complete proteome.  
 SQ SEQUENCE 622 AA; 68093 MM; BE681C01BEA5EB4 CRC64;

Query Match 63.5%; Score 2042; DB 16; Length 622;  
 Best Local Similarity 62.0%; Pred. No. 2.5e-140;  
 Matches 363; Conservative 105; Mismatches 126; Indels 4; Gaps 3;

5 TDYPLAKNHTPDIALSKDQOGLADVRGLTHTVTSISGHPAAGTVELTAAHY 64

DB 6 SDFVLAQANTPNEKQLPOLLPOLADELEFLINKSGSSGHFASGLQVELTAAHY 65  
 QY 65 VFTNPVDQVWDVGHQAYPHKLLTGKREMPTRITLGVSAPAPADESEYDAFGSST 124  
 DB 66 VFTNPVDQVWDVGHQAYPHKLLTGKREMPTRITLGVSAPAPADESEYDAFGSST 125  
 QY 125 SISAALMALASQURGEDEKMAA1IGDGSITGMAVAYEAMNADVANNLVIINDMS1 184  
 DB 126 SISAALMALASQURGEDEKMAA1IGDGSITGMAVAYEAMNADVANNLVIINDMS1 185  
 QY 185 SPPVGMANTLTLYLSKFTSSVRESKRLAKMPVWEIARTEEHKMGTVRG 244  
 DB 186 SPPVGMANTLTLYLSKFTSSVRESKRLAKMPVWEIARTEEHKMGTVRG 245  
 QY 245 LGFRYFPIIDHGVEMLVSTLENLKDLTGPFVHVYTKKGGYAPAEKDPAYHGPAFD 304  
 DB 246 LGFRYFPIIDHGVEMLVSTLENLKDLTGPFVHVYTKKGGYAPAEKDPAYHGPAFD 305  
 QY 305 PTKDFLPKAPSPHPTTTEVFGKWLCDMAADDERLLGTPAMRESGGLVPSQKF 363  
 DB 306 PTKDFLPKAPSPHPTTTEVFGKWLCDMAADDERLLGTPAMRESGGLVPSQKF 363  
 QY 364 FDVALAEOHATVLAAGACOGAKPVVAIYSTFLORGVDLIHDVALQNLMLFALD 423  
 DB 364 FDVALAEOHATVLAAGACOGAKPVVAIYSTFLORGVDLIHDVALQNLMLFALD 423  
 QY 424 VGGPDGPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 482  
 DB 424 VGGPDGPTHAGAPDVSVMCIENMLIMADENECRQMLTTFQHHGPASVYPRGK 483  
 QY 483 AALPPTLTALTEIGABVRHHSRIILAMGSMVTPAVEAGKOLGATVNNRFPVPOAL 542  
 DB 484 AALPPTLTALTEIGABVRHHSRIILAMGSMVTPAVEAGKOLGATVNNRFPVPOAL 543  
 QY 543 VIELARTHDVFTVEENVVLAGAGSAINTELQAKVLMPCVNIQLPDRFVEGSEEE 602  
 DB 544 VIELARTHDVFTVEENVVLAGAGSAINTELQAKVLMPCVNIQLPDRFVEGSEEE 603  
 QY 603 LVGLDSKGLATTEQFCA 620  
 DB 604 LVGLDSKGLATTEQFCA 621

RESULT 3  
 QY 08EGR9 PRELIMINARY; PRT; 620 AA.  
 DB 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 GN DKS OR C0531.  
 OS Bacterioides coli 06.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Bacterioides.  
 OC NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:HI / CPT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 Mobley H.L.T., Donnenberg M.S., Blattner F.R.,  
 "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 RL EMBL; AB016756; AAN79009.1; -.  
 DR TIGR; S01525; -.  
 KW Complete proteome.  
 SQ SEQUENCE 620 AA; 67633 MM; D9DCB76CA65BA7BA CRC64;

Query Match 62.2%; Score 1999.5; DB 16; Length 620;  
 Best Local Similarity 60.0%; Pred. No. 3.2e-137;



Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL: AE014353; AAN29379.1.  
 DR TIGR: BR0436.  
 KW Complete proteome.  
 SQ SEQUENCE 644 AA; 69179 MW; 342DFB4422D129FEA CRC64;

Query Match 49.5%; Score 1592.5; DB 16; length 643;  
 Best Local Similarity 51.4%; Pred. No. 1.7e-107;  
 Matches 318; Conservative 107; Mismatches 181; Indels 13; Gaps 7;

```

OY 8 PLKNIHTPADIRALSKOOLQOLADVRGYLTHVTSISGHPAAGTVEITVLAHYEN 67
DB 7 PLIDRAFPDRLRALPEODLPQLAEELTELIDAVSTTGSHLGLVVELTVLAHYEN 66
OY 68 TPVDOLVMDVGHQAVPHKILTGKREKMPITRTIGVSAFAPADESEYDAFGVSHSSTIS 127
DB 67 TPYDILIMDVGHQAVPHKILTGKRRIRITLQAGLSGFTGPAESEYDPFGAASHSTIS 126
OY 128 AALGMAIASQLKGEDEKKNVAIIGDSITGGMAIEMANHAGDVNANLVLINDNKSISP 187
DB 127 AGLGMAIVASELSEKKNVAIIGDSMSKGAAYEMANNAGLDAPLIVILINDNKSIAIP 186
OY 188 VGAMNNVYLTKVLSKPYSSVRESKKALAKMPS-VWELARKTEBRVKGAVIPGTLFEELG 246
DB 187 TGAMSVYARLVSGRTYRSVRBAKQVAKLPKPTQDPARKSEYARAFPTGTLFEELG 246
OY 247 FNYFPGIDGHVEMLVSTLENLKD-LTGPVFLAVVTKKGGYAPAEKDPFLATRGVAPADP 305
DB 247 FYYVGPIDGHNDHLLPVTKNVRDTQKGPVLIHVVTQKGGYAPAAADAKHGYNKPEV 306
OY 306 TMDPLPKAPSPHPPTTYTVEGKMLCDMAODERLIGTPAMSGSGVSEFSGKPNRPF 365
DB 307 ITGKQAK-PRANVSTYKIFGTSLTBRAHDKIVAVTAAKPTGGLDGEAFKRYVFD 365
OY 366 VALAEQHAVTLAAGACQACAKPVVAIYSTELQRGVDLIHVAALQNLDMLEPALDAGLVG 425
DB 366 VGIAEQHATTPAAGLASSEYKPCALYSTELQRGVDQVHVDSIQNLVPRFPIDRAGLVG 425
OY 426 PGGPTAGAFDYSYKRCINMLIMAPADENECQMLTGFQ-HGPASRYTRGKGPAA 484
DB 426 ADGPPTAGFDTGFLAALPGFVMAASDAEHLHMKVTAEDBGIISFRYRGDGVGD 485
OY 485 IDPTLTALIGKAEVHHGSRITAILAMGSMTPPAVEAKQOLAT-----TVNMRPVKFPD 539
DB 486 LBERGSVLITGKRIYRSGTKVALLSFRIDECILAAAEILTAAGLSTVPAARAKPLD 545
OY 540 QALVLELARTHDFVTVBENVTAGAGSAINFLQAKVLA---MPCNIGLPDRFVBOGS 596
DB 546 HDILRLAREHEVLAVVEEGAV-GGFGSHVLQFLATDGLDRGLKVFALTLPDIYQDHGK 604
OY 597 REELSLVGLDSKGLIATI 615
DB 605 PDAMVAEAGLDRTGIVRTV 623

```

RESULT 6  
 O8KF19 PRELIMINARY; PRT; 644 AA.  
 AC O8KF19, 01-OCT-2002 (TREMELrel. 22, Created)  
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE 1-deoxyxylulose-5-phosphate synthase.  
 GN DKS OR CT0337.  
 OS Chlorobium tepidum.  
 OC Bacteri; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
 OC Chlorobium.  
 NCBI Taxid=1097;  
 OX (1)  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-TLS / ATCC 49652 / DSM 12025;  
 RX MEDLINE=22103685; PubMed=12093901;  
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.

DR Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,  
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
 RA Nieman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,  
 RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,  
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M., TLS, a  
 RT "The complete genome sequence of *Chlorobium tepidum*." a  
 RT photosynthetic, anaerobic, green-sulfur bacterium." a  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
 DR EMBL: AE012812; AAM71583.1;  
 DR TIGR: CT0337;  
 DR InterPro: IPR005477; Dxs.  
 DR InterPro: IPR005476; Transketolase\_C.  
 DR InterPro: IPR005475; Transketolase\_CR.  
 DR InterPro: IPR005474; Transketolase\_N.  
 DR Pfam: PF02780; transketolase\_C\_1.  
 DR Pfam: PF02779; transketolase\_P\_1.  
 DR TIGRFAMs: TIGR00204; dks; 1.  
 DR PROSITE: PS00801; TRANSKETOLASE\_1; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 644 AA; 69739 MW; DE2FB8181AP2B63 CRC64;

Query Match 46.6%; Score 1498.5; DB 16; length 644;  
 Best Local Similarity 49.3%; Pred. No. 1.3e-100;  
 Matches 308; Conservative 93; Mismatches 211; Indels 13; Gaps 6;

```

OY 7 YPLKNIHTPADIRALSKOOLQOLADVRGYLTHVTSISGHPAAGTVEITVLAHYE 66
DB 15 YPLLSHSHPADLKQSLHLELVAACRKVLELSQNGHGSSGAVVELTVLAHYE 74
OY 67 NTPVDOLVMDVGHQAVPHKILTGKREKMPITRTIGVSAFAPADESEYDAFGVSHSSTI 126
DB 75 QSPDRIIMDVGHQAVPHKILTGKRRIRITLQAGLSGFTGPAESEYDPFGAASHSTI 134
OY 127 SAALGMAIASQLKGEDEKKNVAIIGDSITGGMAIEMANHAGDVNANLVLINDNKSISP 186
DB 135 SAAAGLAAARLALARKCKKVAIIGDSITGGMAIEMANHAGDVNANLVLINDNKSISP 194
OY 187 PVGAMNNVYLTKVLSKPYSSVRESKKALAKMPS-VWELAR-----KTEBHVKGAVIPGTL 241
DB 195 STGGKKNVNLVTLAKTYNLRKFWPMSLSLHNEIGETAKTAVHREDGILKAFTPGAY 254
OY 242 FEELGFNYFPGIDGHVEMLVSTLENLKD-LTGPVFLAVVTKKGGYAPAEKDPFLATRGV 300
DB 245 FEALGFRYFGPIDGHNDHLLPVTKNVRDTQKGPVLIHVVTQKGGYAPAAADAKHGYNKPE 314
OY 301 PAPPTDPLPKAPSP-HPPTTYTVEGKMLCDMAODERLIGTPAMSGSGVSEFSGKPNRPF 359
DB 315 GGFIDETGKNVKAQKPKPTQEVFGALVELLKDPTTITATTAAPSGTSLDFQAI 374
OY 360 PNRYFDVALAEQHAVTLAAGACQACAKPVVAIYSTELQRGVDLIHVAALQNLDMLEPALD 419
DB 375 PSRCDDVGLAEQHAVTPAAGLASSEYKPCALYSTELQRGVDQVHVDSIQNLVPRFPIDRAGLV 434
OY 420 RAGLVGPDPPTAGAFDYSYKRCINMLIMAPADENECQMLTGFQ-HGPASRYTRGKGPAA 478
DB 435 RAGLVGEDGPTAGAFDLSYLVNVPNLITAPAGDQELRNLTALYDILKSPVALIRPG 494
OY 479 KQPGALIDPTLTALIGKAEVHHGSRITAILAMGSMTPPAVEAKQOLAT-----TVNMR 533
DB 495 SGSGATLHKETPLPVGVGRILRDGKSVALLGITSMSBALBTALAEAGLDPVCDMR 554
OY 534 TVKPPDQALVLELARTHDFVTVBENVTAGAGSAINFLQAKVLAFCVCHIGLPDRVE 593
DB 555 FLKPLDTEIILMAASRCHTIYVITENSITGGFGSNVNYLHHAHFGIKCSIFGLDAEVT 614
OY 594 QGSRREILSLVGLDSKGLIATIRPF 618
DB 615 HGSMDLIRVGLDARSISGLIEF 639

```

RESULT 7  
 082676

ID 082676 PRELIMINARY; PRT; 716 AA.  
 AC 082676;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE 1-deoxy-xylose 5-phosphate synthase.  
 OS Catharantus roseus (Rosy periwinkle) (Madagascar periwinkle).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; Lamiales; Gentianales; Apocynaceae; Rauvolfiaceae; Vinaceae;  
 OC Catharantus.  
 NCBI\_TaxID=4058;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chated K., Oudin A., Guivarc'h N., Hamdi S., Chentleux J.C., Rideau M.,  
 RA Clastre M.;  
 RT "1-deoxy-D-xylose 5-phosphate synthase from periwinkle: cDNA  
 RT identification and induced gene expression in terpenoidindole  
 RT alkaloid-producing cells.";  
 RL Plant Physiol. Biochem. 38:559-566 (2000).  
 DR EMBL; AJ011840; CAA09804.2; -.  
 DR InterPro; IPR005477; Dxs.  
 DR InterPro; IPR005476; Transketolase\_C.  
 DR InterPro; IPR005475; Transketolase\_CR.  
 DR Pfam; PF02780; transketolase\_C\_1.  
 DR TIGRFAMs; TIGR00204; dxs; 1.  
 DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.  
 DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.  
 SQ SEQUENCE 716 AA; 76793 MW; E5D75ED703246F7D CRC64;

Query Match 46.1%; Score 1482.5; DB 10; Length 716;  
 Best Local Similarity 48.3%; Pred. No. 2.2e-99;  
 Matches 307; Conservative 109; Mismatches 188; Indels 31; Gaps 11;  
 8 PLKNIHFPADIRALSKDLOQLADVRGYLTHTVVISGSHFAAGTVELTVALHYEPTP  
 72 ILIDITINPVHMKNTSADHLEQLAEIRAEIYSAKVGSHLSAGVVDLTVALHVFN 131  
 68 TPVDOLWVDVGHQAVPHKILTRKESRMPITRTLGVSAPPADESEYDAFGVGHSSIS 127  
 132 TPEDHIIWVGHQAVPHKILTRKESRMPITRTLGVSAPPADESEYDAFGVGHSSIS 191  
 128 ALGALIASQLGEGDKKVAIIIGDSITGMAVEAMNNAHDVNAHLYLINDNDSISP 186  
 192 AGAGAAVADIIIGKNNVSVIGDGAHTAGVYEAAMNNAHFLDNLIVLINDNKVSLPT 251  
 187 -----PVGANNVLTLYVSSKPYSSVRESKKALAKM--PSWELARKTEEHVKMI- 236  
 252 ATLDDGATFVAGLSALSKIQSPKREARAKSITKQIGFAHEVAAKVDEYARGLS 311  
 237 -VPGITFELGNNYRGPIDGHVEMLVSTLENLXOL--TGPFVLTWTKKGGVAPAKD 293  
 312 ATGSTLFEELGIYYIGPVDSHIEDLVITFYKAKAPAPGPAVLHIVKGGPPAPABA 371  
 294 PLAYHGVAPADPT--KDFLPKAAPEPHPTVEFGRLCDMAADERLIGITPAMEBSG 351  
 372 ADRMGVAVFDPKTKGQFSK---SPTLSYTFYASLLKEADIMKIIAIHMAAGCGTG 428  
 352 LVEFSQKPNRYFDVALAEQHAVTLAAGACOGAKPVVAIYSTFLQRTGDIHVALON 411  
 429 LNVFQKFPDRCGDVIAEOHVTFAAGLATEGLKPCFAIYSSFQROGDVVHVDLQK 488  
 412 LDMFLALDAGLVGPGPTHAGAFDYSYRCIPNNLIMAPADENECRQMLTTGPF--HBP 470  
 489 LPRPAMDAGLVGADGPTHGAFDVAVALCLPMIIVAPSDBAELMHVAVAPAKIDRP 548  
 471 ASVRYRGKGPALDPTL--TALIGKAEVRHHSRIIILAMGSMVTPAVEAKOL--- 525  
 549 CCRFPFRGNGIGVALPPNKGTPLEIGKRIIVESGRVAHIGVGSIVQOCLGAAMELKS 608  
 526 --GATVNNREFYKPEFOCALVLELARTHDVYVTEENVVIAAGASALNTEFQAKVL--M 580  
 609 NVSPVADMAFCPLDGDILKTLKEHEILITVEBSI--GGFSGVTHFLSLTGLIDGI 667

Query 581 PNCNIGPDRFEVQSGREBLISVGLDSKGLIATI 615  
 668 KVASLFLPDRYIDHGAVPDQIEAGLSRRHICATI 702

## RESULT 8

ID 08DL74 PRELIMINARY; PRT; 655 AA.  
 AC 08DL74;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE 1-deoxy-xylose 5-phosphate synthase.  
 OS Synecococcus elongatus (Thermosynechococcus elongatus).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
 NCBI\_TaxID=32046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STEALIN=BP-1;  
 RX MEDLINE=2225144; PubMed=12240834;  
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
 RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
 RA Shingo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,  
 RT "Complete genome structure of the thermophilic cyanobacterium  
 RT Thermosynechococcus elongatus BP-1.";  
 RL DNA Res. 9:123-130 (2002).  
 DR EMBL; AP005371; BAC08174.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 655 AA; 70984 MW; 36E5ABABF89A2D8 CRC64;

Query Match 45.2%; Score 1455; DB 16; Length 655;  
 Best Local Similarity 47.8%; Pred. No. 1.9e-97;  
 Matches 298; Conservative 108; Mismatches 201; Indels 16; Gaps 8;  
 10 LKNHTPADIRALSKDLOQLADVRGYLTHTVVISGSHFAAGTVELTVALHYEPTP 69  
 20 LSELHNNVQAGLSIAQLTQIAQIRKHEIYVAATGHLGPGVVELTVALVOTDLE 79  
 70 VDOLWVDVGHQAVPHKILTRKESRMPITRTLGVSAPPADESEYDAFGVGHSSISAA 129  
 80 KDRVWVDVGHQAVPHKILTRKESRMPITRTLGVSAPPADESEYDAFGVGHSSISAA 139  
 130 LGALIASQLGEGDKKVAIIIGDSITGMAVEAMNNAHDVNAHLYLINDNDSISP 188  
 140 LGALIASQLGEGDKKVAIIIGDSITGMAVEAMNNAHDVNAHLYLINDNDSISP 199  
 189 GANNVLTLYVSSKPYSSVRESKKALAKMPSV---LAKTEEHVKMIVP--GTLF 242  
 200 GAIRYINKRLSPQVQFPTDNLEQFPHIIPFGENLTPEKORLKEGKRLAVPKGAVF 259  
 243 BELGNTFPGPIDGHVEMLVSTLENLXOLGPFVLTWTKKGGVAPAKDPLAYHGVPA 302  
 260 BELGNTFPGPIDGHVEMLVSTLENLXOLGPFVLTWTKKGGVAPAKDPLAYHGVPA 319  
 303 PD-PTKDFLPKAAPEPHPTVEFGRLCDMAADERLIGITPAMEBSG 361  
 320 FDLVTGAKESKRP--PSYKVFGLTTLKAEVDPIVIGITAMAGTGLDILQKVPK 378  
 352 RYFDVALAEQHAVTLAAGACOGAKPVVAIYSTFLQRTGDIHVALONLDMFLADRA 421  
 379 QYIDVGLAEQHAVTLAAGACOGAKPVVAIYSTFLQRTGDIHVALONLDMFLADRA 438  
 422 GLVPGDPTHAGAFDYSYRCIPNNLIMAPADENECRQMLTTGPF--HBPASVYPRKSG 480  
 439 GLVPGDPTHAGAFDYSYRCIPNNLIMAPADENECRQMLTTGPF--HBPASVYPRKSG 498  
 481 PCNAL-DPTLALTEIGKAEVRHHSRIIILAMGSMVTPAVEAKOL---GATVNNREF 534  
 499 YGVALMEGMEPLIEIGKELLSGEBDILVAAGSMVVPAMGVAILKHEGMSAAVINRF 558

```

QY 535 VKPFDQALVELARFTHDVFTVEENVIAGGASALNTFLAQKYLMPVNCIGLDPREVEQ 594
DB 559 AKPFDLTLPLAQLQGRVFTLEEGCGAGGSAVLENLQADILVFLKGLVDFLVER 618
QY 595 GSREELSLVGLDSKGLIATIEQ 617
DB 619 ASPDESKADIGLTPPQMAETIMQ 641

RESULT 9
Q8L692 PRELIMINARY: PRF; 711 AA.
AC Q8L692;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE 1-deoxy-D-xyulose 5-phosphate synthase 2 precursor,
GN DXS2.
OS Medicago truncatula (Barrel medic.).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
OC NCBI_TaxID=3880;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root;
RA Walter M.H., Hans J., Strack D.;
RT "Two distantly-related genes encoding 1-deoxy-D-xyulose 5-phosphate
RT synthases: differential regulation in shoots and apocoteneoid-
RT accumulating mycorrhizal roots."
RL Submitted (FBS-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430048; CAD22531.1; -
DR InterPro; IPR005477; Dxs.
DR InterPro; IPR005476; Transketolase_C.
DR InterPro; IPR005475; Transketolase_CR.
DR InterPro; IPR005474; Transketolase_N.
DR Pfam; PF02780; transketolase_C_1.
DR Pfam; PF02779; transketolase_Pyr_1.
DR TIGRFAMs; TIGR00204; dxs; 1.
DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
KW Signal.
KW CHAIN.
FT SIGNAL 1 44 POTENTIAL.
FT CHAIN 45 711 1-DEOXY-D-XYULOSE 5-PHOSPHATE SYNTHASE.
SQ SEQUENCE 711 AA; 76940 MW; E920A2B8E2F8C9A CRC64;

Query Match
Best Local Similarity 44.9%; Score 1444.5; DB 10; Length 711;
Matches 302; Conservative 112; Mismatches 192; Indels 27; Gaps 10;

QY 8 PLKKNITPADIRALSQOLQOLADENVGYLTHVTSISGGHFAAGLVEELVLAHYVN 67
DB 67 PLIDTVPFVPMKMLTTEDLEQLAELRADIVHSVDTGHLSSLSGLVEELVALHVF 126
QY 68 TPVDQLVWDVGHQAYPHKILITGRKEMPTITRTLGVSAPFAPADSEEDAFVGGHSTIS 127
DB 127 TDDDKIMVGHQAYPHKILITGRSRMHTIRKTSGLAGPKKDESVDAFVGGHSTIS 186
QY 128 AALGAVALSOLGEGDKKVAIIIGDSITGCAAYEAMNHAADVANNLLVILINDMSISP- 186
DB 187 AGIGMAVAVADLLGKONSVISVIGDGMATAGQAYEAMNNAAGFIDSLIYLINNNKOVSLPT 246
QY 187 -----PVGANNVYLTIVLSKSFYSVRESKALAKAPSWML-ARKTEEHVKGMI- 236
DB 247 ATLDPATVPGALSLKSLQASKEFKRLKREATKNTITKQIGQTHLVASKVDKARDFIS 306
QY 237 -VPGTLFEELGNYFGPIDGHDVEMLVSTLENKDL--TGYPLAVYTKKGGYAPAKED 293
DB 307 GSGSSLFEEELGMYIIGPMGNHNIIDLVINIFEKYKAMPAPGPVLIHIVTEKGGYEPALAA 366
QY 294 PLAYGVVPAFDPDKQFLPKAAPSPHPTVEVFGNMLCDMAADBERLGLTPMAREGSLV 353
DB 367 ADRMGGVVKFDPKTHQFKPKDST-LATVYQFADSLIKELKENDKIKVALHAAANGGTGLN 425

```

```

QY 354 EFSQKZPNRYFVPAIAEQAVALTAAQACQAKPVVAIVSTFLQSGDQLIHDAVONLD 413
DB 426 YQKRPDPCFVQVIAEQAVALTAAQACQAKPVVAIVSTFLQSGDQLIHDAVONLD 485
QY 414 MFLADRAGLVGGDGTGAGADYSYMRCTPNNLMAADDEVCROMLTGGO-HHGPA 472
DB 486 VRFAMDRAGLVGADGPTHGADDTFMACTPNNIVAPSDDEALMMVATAAIDDRPSC 545
QY 473 VRRPGRKGQA--AIDPTLTALIEGAEVRRHSGRIATILAMGSMVTPAVEAGKQGA-- 527
DB 546 FRPFRGNGIGANLPLNNKGTPIEIGGRILLBSRRVAILIGYCMYQCKKAAEMLAAYG 605
QY 528 --TVNNRPFVKPFDQALVELARFTHDVFTVEENVIAGGASALNTFLAQKYL---MPV 582
DB 606 YVTVADARFCKPDLTDLIRLAREHILITVEEGSI-GGGRSHVSQFLSLAGLDGFLKL 664
QY 583 CNIGLADRPVQSGREELSLVGLDSKGLIAT 615
DB 665 RSMMLPDRITDGAQPDQIDRGLSKHILATV 637

```

## RESULT 10

Q8L6C6 PRELIMINARY: PRF; 715 AA.

```

AC Q8L6C6;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
DE 1-deoxy-D-xyulose 5-phosphate synthase.
GN DXS.
OS Stevia rebaudiana (Stevia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Eupatoriaceae; Stevia.
OC NCBI_TaxID=55670;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue N.M.L.C.;
RA "Cloning and expression of cDNAs encoding two enzymes of the MEP
RT pathway in Stevia rebaudiana Bertoni."
RL Plant Physiol. 0:0-0(0).
DR EMBL; AJ429232; CAD22155.2; -
DR InterPro; IPR005477; Dxs.
DR InterPro; IPR005476; Transketolase_C.
DR InterPro; IPR005475; Transketolase_CR.
DR Pfam; PF02780; transketolase_C_1.
DR Pfam; PF02779; transketolase_Pyr_1.
DR TIGRFAMs; TIGR00204; dxs; 1.
DR PROSITE; PS00802; TRANSKETOLASE_2; 1.
DR PROSITE; PS00801; TRANSKETOLASE_1; 1.
SQ SEQUENCE 715 AA; 76559 MW; A6C430C6FA2AFEBB CRC64;

```

Query Match  
Best Local Similarity 44.8%; Score 1440.5; DB 10; Length 715;  
Matches 303; Conservative 108; Mismatches 193; Indels 31; Gaps 12;

```

QY 8 PLKKNITPADIRALSQOLQOLADENVGYLTHVTSISGGHFAAGLVEELVLAHYVN 67
DB 67 PLIDTVPFVPMKMLTTEDLEQLAELRADIVHSVDTGHLSSLSGLVEELVALHVF 126
QY 68 TPVDQLVWDVGHQAYPHKILITGRKEMPTITRTLGVSAPFAPADSEEDAFVGGHSTIS 127
DB 68 TPDDKIMVGHQAYPHKILITGRSRMHTIRKTSGLAGPKKDESVDAFVGGHSTIS 186
QY 128 AALGAVALSOLGEGDKKVAIIIGDSITGCAAYEAMNHAADVANNLLVILINDMSISP- 186
DB 128 TPDDKIMVGHQAYPHKILITGRSRMHTIRKTSGLAGPKKDESVDAFVGGHSTIS 187
QY 128 -----PVGANNVYLTIVLSKSFYSVRESKALAKAPSWML-ARKTEEHVKGMI- 236
DB 247 ATLDPATVPGALSLKSLQASKEFKRLKREATKNTITKQIGQTHLVASKVDKARDFIS 306
QY 237 -VPGTLFEELGNYFGPIDGHDVEMLVSTLENKDL--TGYPLAVYTKKGGYAPAKED 293
DB 186 AGIGMAVAVADLLGKONSVISVIGDGMATAGQAYEAMNNAAGFIDSLIYLINNNKOVSLPT 247
QY 187 -----PVGANNVYLTIVLSKSFYSVRESKALAKAPSWML-ARKTEEHVKGMI- 237
DB 248 ATLDPATVPGALSLKSLQASKEFKRLKREATKNTITKQIGQTHLVASKVDKARDFIS 307

```

QY 238 P--GTLFEELGRTYRPIGDHDEMLVSTLENLKD--TGPPVLAHVYTKKGGYAPAEKD 293  
 DB 308 AGSSTLFEELGRTYRPIGDHDEMLVSTLENLKD--TGPPVLAHVYTKKGGYAPAEKD 367  
 QY 294 PLAYHGVPAFD-PT-KDPLPKAASPHPTTEVFGWMLCDMAAODERLIGTPARESGS 351  
 DB 368 ABRMGVAVFDPYTKQFKTK--SPTLSYTOYFASLKEBAEDNKVLAITANAGGTTG 424  
 QY 352 LVEFSQKPNRYFDVAIAQAVTLAAGACQAGKPVAVIYSTPFGYDQILHVALON 411  
 DB 425 LNYFQKCEPCERCFDVGIAQHAHTFAAGLATEGLKPFCAIYSSFLQRYDYQVHVDLQK 484  
 QY 412 LMLFLDAGLVGDPGPTTHAGAFDYSYRCCI PNMLIMPADENECROWLTGFGH-HNEP 470  
 DB 485 LVRFPMDRAGLVGAGPPTHGAPDITTYACLPNNVWVAPDAEIMHVAATAAALIDRRP 544  
 QY 471 ASVRYPRGKPGCAIDPTLTALA--LEIGKARVHHSRIAILAMGSMVTPAVEAKOL--- 525  
 DB 545 SCRRPFRNGIGAPLEPNKKGPIEVGKGRILLEGTRVAILGYGSIYVGCAGASILQAH 604  
 QY 526 --GATVMMRFPKPPQALVBLARHDPVYVEENVVAGGASINTLOAKVL---M 580  
 DB 605 NVSATVADARFCPPDPTGLIRLANEHEVLLTYEBGSI-GGFGSHVAFSLINGLIDKTL 663  
 QY 581 PVCNIGLPDRFVEQSRREELSLVSGDSKGIATI 615  
 DB 664 KLRAMTLPRKTIIDKGAPODQLEBTGLSSTHICSSL 698

## RESULT 11

Q8K213 PRELIMINARY; PRT; 630 AA.

AC Q8K213;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Deoxyxylylulose-5-phosphate synthase.  
 GN DXS.  
 OS uncultured proteobacterium.  
 OC Bacteria; Proteobacteria; environmental samples.  
 NC NCB1\_TaxID=153809;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2182632; PubMed=11832943;  
 RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,  
 Hamada T., Eisen J.A., Fraser C.M., DeLong E.P.,  
 RT "Unuspected diversity among marine aerobic anoxygenic phototrophs.";  
 RL Nature 415:630-633 (2002).  
 DR EMBL; AE008921; AAM48660.1; -;  
 DR InterPro; IPR005477; Dxs.  
 DR InterPro; IPR005476; Transketolase C.  
 DR InterPro; IPR005475; Transketolase CR.  
 DR InterPro; IPR005474; Transketolase N.  
 DR Pfam; PF02780; transketolase\_C.1.  
 DR Pfam; PF02779; transket. pyr; 1.  
 DR TIGRFAMs; TIGR0204; dxs; 1.  
 DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.  
 DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.  
 SQ SEQUENCE 630 AA; 65840 MW; 26789DADF48111C97 CRC64;

Query Match 44.7%; Score 1437.5; DB 2; Length 630;  
 Best Local Similarity 47.6%; Pred. No. 3 4e-96;  
 Matches 302; Conservative 100; Mismatches 195; Indels 37; Gaps 9;

QY 7 YPLKNHTPADIRALSKDQLOQLADEVRGYLTHVTSISGGHFAAGLGTVELTVALHYVF 66  
 DB 2 HSLLEITLPSDLTKTSLDQQLADELRVDVIEANSRFGHSGSLGVEVETVALHVF 61  
 QY 67 NRPVQQLVWDVGHQAVPHKILNGRKEKMPITRELIGSVSAFPAADSEVYAPGVGSSSTSI 126  
 DB 62 DHPKDKLMDVGHQCPHAKVYNGRRKMGKTLROGGSLGFTKSSSEVPYPPAAASSSTSI 121  
 QY 127 SAALGMAIASQLRGEDKKNVAIIGDGSITGNAYEAMNAGDVNKLVLINDMDSISP 186

DB 122 SAHGFVARDLGGDTGDALIVIDGSIAGMAVEALNAGABGRFPVILLNDKSTAP 181  
 QY 187 PVGAMNNYITFYLLSKSYSSVREESKALAMPSS-----VWELARKTEEHYXG 234  
 DB 182 PVGAMSKMSGLAGI-----ALAQLMFGQDIETVLDPQPRDHRARRARELVATG 229  
 QY 225 MIV-PGTLFEELGRTYRPIGDHDEMLVSTLENLKD--LTGPPVLAHVYTKKGGYAPAEK 292  
 DB 220 AVASGQITLFEELGRTYRPIGDHDEMLVSTLENLKD--LTGPPVLAHVYTKKGGYAPAEK 289  
 QY 203 DPLAYHGVPAFDPTKDFPKAASPHPTTEVFGWMLCDMAAODERLIGTPARESGSL 352  
 DB 290 SADKTHGVAKPDVAGAGMGSGANR-PSTYTFPGTLTKLAKOSKIVGITAAHPSGGL 348  
 QY 353 VBSQKPNRYFDVAIAQAVTLAAGACQAGKPVAVIYSTPFGYDQILHVALON 412  
 DB 349 DIFAPRFDMEHVDGIAEQHGVTFAGAGLGFPCAIYSTPFGYDQILHVALON 408  
 QY 413 DMLFLDAGLVGDPGPTTHAGAFDYSYRCCI PNMLIMPADENECROWLTGFGH-HNEP 471  
 DB 409 PVRPAIDAGLVGAGPPTHGAPDITTYACLPNNVWVAPDAEIMHVAATAAALIDRRP 544  
 QY 472 SVRYPRGKPGCAIDPTLTALA--LEIGKARVHHSRIAILAMGSMVTPAVEA-----GKOLG 526  
 DB 469 AFRYPRGTGVSIPQEGELLQIGKRIVRGAEIALLSFGHLAEALKAADLISAQGV 528  
 QY 527 ATVMMRFPKPPQALVBLARHDPVYVEENVVAGGASINTLOAKVL---M 580  
 DB 529 ATVADARFAPKPLDHALISKLAKTHKVLITTEOGA-QGFGAMVLAHYLADTGLDSPLVR 587  
 QY 584 NIGLPDRFVEQSRREELSLVSGDSKGIATI 617  
 DB 588 SMTLPDRITDQAPADPAMYADAGLATDIAATLQ 621

## RESULT 12

Q9FSG2 PRELIMINARY; PRT; 709 AA.

AC Q9FSG2;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE 1-D-desoxyxylylulose 5-phosphate synthase (DXS).  
 GN DXS.  
 OS Narcissus pseudonarcissus (Daffodil).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;  
 NC Narcissus.  
 OC NCB1\_TaxID=39639;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Schaub P., Beyer P., Al-Badli S.;  
 RT "A cDNA encoding 1-D-desoxyxylylulose 5-phosphate synthase (DXS) from  
 RL Narcissus pseudonarcissus L.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ279019; CAC08458.1; -;  
 DR InterPro; IPR005477; Dxs.  
 DR InterPro; IPR005476; Transketolase C.  
 DR InterPro; IPR005475; Transketolase CR.  
 DR InterPro; IPR005474; Transketolase N.  
 DR Pfam; PF02780; transketolase\_C.1.  
 DR Pfam; PF02779; transket. pyr; 1.  
 DR TIGRFAMs; TIGR0204; dxs; 1.  
 DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.  
 DR PROSITE; PS00802; TRANSKETOLASE\_2; 1.  
 SQ SEQUENCE 709 AA; 75747 MW; 1FBA362A0DEDC6D3 CRC64;

Query Match 44.6%; Score 1435.5; DB 10; Length 709;  
 Best Local Similarity 47.6%; Pred. No. 5 8e-96;  
 Matches 302; Conservative 107; Mismatches 195; Indels 31; Gaps 11;

QY 8 PLKNHTPADIRALSKDQLOQLADEVRGYLTHVTSISGGHFAAGLGTVELTVALHYVF 67



DB 65 PLDTITVPDPAHKNLSSTODLEBOLASRLDITVSVSKTGGHISASLGVNDLTVLHVPD 124

QY 68 TPVDQVWDVGHQAYPHKILITGRKERMPTITRTIGVSAFPADESEYDAFGVSHSTIS 127  
 DB 125 TDDDKIMDVGHQAYPHKILITGRSRMHTLRQTSGLAGFPKDESDVDAFGVSHSTIS 184

QY 128 AALGMAIASQLRGSDKQVAIIIGDSTITGGMAIYEMNAGDVANALVILNDNISIP- 186  
 DB 185 AGLGMAVAGRLGKGTNNVVAIIIGDGMATQAYEMNAGDVANALVILNDNQVSLPT 244

QY 187 -----PVGAMNNYLTKLSSKFYSVREESKALAKM-PSWELARKTEEHVKMI 237  
 DB 245 AALNGPAPVGHLSGLALQASAKFROLRKAKICITQIGGQAEVAAKDEYARGNIS 304

QY 238 P--GTLFEEELGPNYFPIGDHDEMVLSTLENKDL--TGPFVFLAVVTKKGGYAPAEKD 293  
 DB 305 ASGASLFEELGLYYIGVYDGHVNEBVLVAFKKYKMPSPGPAVVAIVTEKGGYPPAARA 364

QY 294 PLAYHGVPAHPPT--KDFLRKAAPSPHTYTEVGRMLCDMAAGDELLGITPAMRESG 351  
 DB 365 ADKTHGVGRKDPATGKPFKSA--PTQSTYTPFAELIABEADKCIYGIHAAAGGSGT 424

QY 352 LVEFSQKFPNRYFDVAIAEOHVAVTLAAGQACQAKPVAIYSTFLQGYDQILHDAVLON 411  
 DB 422 LNFQKKEFPDRCFVGIABQHAVTFAAGLATEGLKPCFCAIYSSFLQGYDQVHVDLQK 481

QY 412 LDMFLALDPRAGVDPDPTAGAFDYSYRCIPNNLIMAPDEBCROMLTTGQ--HNGP 470  
 DB 482 LVPFRLDPRAGVADPPTAGAFDYSYRCIPNNLIMAPDEBCROMLTTGQ--HNGP 541

QY 471 ASVRYPRGKPGAAI--DPTLALSIGKAEVRHSGRIALIMAGSMYTPAVEAG-----K 523  
 DB 542 SCFRPRGNGIGVLPFGNGVPLEVGRKIMLBQGVALLGYGTAVQSCMAATVOER 601

QY 524 QLGATVVMKRFKFPDQALVLELARTHDVFTVYVENVIAGAGAGINTFLQACVLA--M 580  
 DB 602 GSATVADRFCKPDLSELIRLVNEHEILITVEEGSI--GQFASHVSHFLSLGLDGL 660

QY 581 PVCNIGLPDRFVQGSREBELSLVGLDSKGIAT 615  
 DB 661 KLRSMVLPDRYIDHGAAPKQIEAGLSKGIATV 695

RESULT 13  
 Q9SP65 PRELIMINARY: PRT: 713 AA.

AC Q9SP65: 01-MAY-2000 (TREMREL. 13, Created)  
 DT 01-OCT-2000 (TREMREL. 15, last sequence update)  
 DT 01-OCT-2002 (TREMREL. 22, last annotation update)  
 DE 1-deoxy-D-xylose-5-phosphate synthase.  
 GN DKS1.  
 OS Artemisia annua (Sweet wormwood).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asterales; Campanulids; Asterales; Asterales; Asteroideae;  
 OC Arctemisia; Artemisia.  
 OC NCBI\_TaxID=35608;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Yugoslavia; TISSUE=Root;  
 RA Mober K.X., Souret F.F., Shore K.A., Weathers P.J.;  
 RT "Artemisia annua D-1-deoxyxylose-5-phosphate synthase (dxs) mRNA";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBS databases.  
 DR BMB, AF182286; AAD56390.2; -;  
 DR InterPro; IPR005477; Dxs.  
 DR InterPro; IPR005476; Transketolase\_C.  
 DR InterPro; IPR005475; Transketolase\_CR.  
 DR InterPro; IPR005474; Transketolase\_N.  
 DR Pfam; PF02780; transketolase\_C\_1.  
 DR Pfam; PF02779; transket. pyr. 1.  
 DR TIGRFAM; TIGR00204; dxs; 1.  
 DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.

SO SEQUENCE 713 AA; 76669 MM; 403FP36512C2372A CRC66;  
 Query Match 44.6%; Score 1435.5; DB 10; Length 713;  
 Best Local Similarity 47.5%; Pred. No. 5.9e-96;  
 Matches 303; Conservative 111; Mismatches 187; Indels 37; Gaps 13;

QY 8 PLNKHITPADIRALSDQLOQLADEVRYGLTRHVSISGCFPAAGCTVETVALHVEN 67  
 DB 68 PLDTITVPDPAHKNLSSTODLEBOLASRLDITVSVSKTGGHISASLGVNDLTVLHVPD 124

QY 68 TPVDQVWDVGHQAYPHKILITGRKERMPTITRTIGVSAFPADESEYDAFGVSHSTIS 127  
 DB 125 TDDDKIMDVGHQAYPHKILITGRSRMHTLRQTSGLAGFPKDESDVDAFGVSHSTIS 184

QY 128 AALGMAIASQLRGSDKQVAIIIGDSTITGGMAIYEMNAGDVANALVILNDNISIP- 186  
 DB 185 AGLGMAVAGRLGKGTNNVVAIIIGDGMATQAYEMNAGDVANALVILNDNQVSLPT 244

QY 187 -----PVGAMNNYLTKLSSKFYSVREESKALAKM-PSWELARKTEEHVKMI 237  
 DB 245 AALNGPAPVGHLSGLALQASAKFROLRKAKICITQIGGQAEVAAKDEYARGNIS 304

QY 238 P--GTLFEEELGPNYFPIGDHDEMVLSTLENKDL--TGPFVFLAVVTKKGGYAPAEKD 293  
 DB 305 ASGASLFEELGLYYIGVYDGHVNEBVLVAFKKYKMPSPGPAVVAIVTEKGGYPPAARA 364

QY 294 PLAYHGVPAHPPT--KDFLRKAAPSPHTYTEVGRMLCDMAAGDELLGITPAMRESG 351  
 DB 365 ADKTHGVGRKDPATGKPFKSA--PTQSTYTPFAELIABEADKCIYGIHAAAGGSGT 424

QY 352 LVEFSQKFPNRYFDVAIAEOHVAVTLAAGQACQAKPVAIYSTFLQGYDQILHDAVLON 411  
 DB 422 LNFQKKEFPDRCFVGIABQHAVTFAAGLATEGLKPCFCAIYSSFLQGYDQVHVDLQK 481

QY 412 LDMFLALDPRAGVDPDPTAGAFDYSYRCIPNNLIMAPDEBCROMLTTGQ--HNGP 470  
 DB 482 LVPFRLDPRAGVADPPTAGAFDYSYRCIPNNLIMAPDEBCROMLTTGQ--HNGP 541

QY 471 ASVRYPRGKPGAAI--DPTLALSIGKAEVRHSGRIALIMAGSMYTPAVEAG-----K 523  
 DB 542 SCFRPRGNGIGVLPFGNGVPLEVGRKIMLBQGVALLGYGTAVQSCMAATVOER 601

QY 524 QLGATVVMKRFKFPDQALVLELARTHDVFTVYVENVIAGAGAGINTFLQACVLA--M 580  
 DB 602 GSATVADRFCKPDLSELIRLVNEHEILITVEEGSI--GQFASHVSHFLSLGLDGL 660

QY 581 PVCNIGLPDRFVQGSREBELSLVGLDSKGIAT 615  
 DB 661 KLRSMVLPDRYIDHGAAPKQIEAGLSKGIATV 695

RESULT 14  
 Q8GAA0 PRELIMINARY: PRT: 636 AA.

AC Q8GAA0: 01-MAR-2003 (TREMREL. 23, Created)  
 DT 01-MAR-2003 (TREMREL. 23, last sequence update)  
 DT 01-MAR-2003 (TREMREL. 23, last annotation update)  
 DE 1-deoxy-D-xylose-5-phosphate synthase (EC 4.1.3.37).  
 GN DKS.  
 OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
 OC NCBI\_TaxID=1140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20026086; Pubmed=10556522;  
 RA Miller B., Heuser T., Zimer W.;  
 RT "A Synecococcus leopoliensis SAUG 1402-1 operon harboring the 1-deoxyxylose-5-phosphate synthase gene and two additional open reading frames is functionally involved in the dimethylallyl diphosphate synthesis";  
 RL FBS Lett. 460:485-490(1999).  
 RN [2]



RP SEQUENCE FROM N.A.  
RA Holtman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,  
RA McMurtry S., Gonzalez A., Salinas I., Golden S.S., Youderian P.,  
RT "Synechococcus elongatus PCC 7942 cosmid 389."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RM EMBL: X04616; CDS5546.1; .  
KM Lysase.  
SQ SEQUENCE 636 AA; 69328 MW; 51838C3A5AA7A3D CRC64;  
Query Match 44.6%; Score 1433; DB 2; Length 636;  
Best Local Similarity 48.1%; Pred. No. 7.4e-96;  
Matches 300; Conservative 99; Mismatches 209; Indels 16; Gaps 8;  
QY 10 LKNHTPADIALSKDQIQQLADREYRGLTHTVSISGHPAAGTVELTVAHYVNTVP 69  
DB 3 LSEITHPQAGLSVAQIQGHIQREHQTVAATGHLPGGLGVELTVALVQTLDL 62  
QY 70 VDQVWDVGHQAYPHKILTGKREMPTRITLGSVAFAPADESEYDAFGVSHSTISAA 129  
DB 63 RDKVWVDGHQAYPHKILTGKRYNPHITRQKDLGAGLYKRENFPHFGAGHASTISAA 122  
QY 130 LGMALASQSGEDKKVAIIGDGSITGMAYEAMNAGDV-NANLVLINDNMSISPPV 188  
DB 123 LGMALASQSGEDRCVAVIIGDGSITGMALAEINHGHPKRLVYLANDNMSISPNV 182  
QY 189 GAMNNYLTQVLSKPYSSVREESKKALAKMPVWEAKRTEHYK-GM-----IVPGLF 242  
DB 183 GALSSTYLAKLIVSEPMQLITDGLTQGMQQLPFGALITQGFPEVKEGKRLSYKIGAVP 242  
QY 243 BELGFNYFPIGDHVMVLSTLENLXDLTGPVFLAVVTKKGYAPAEKDPPLAHGVPA 302  
DB 243 ELGFTYVGPVDGNHLELITPBAHNGTGPVLAHTTKGKGYPAEBDQVGHQNP 302  
QY 303 PD-PTKQPLPKAASPHPTTYTEVFGRLCDMAADERLGITTAMRBSGGLVESQKPEP 361  
DB 303 FDLATGAKAPASKRP-PSYKVPQQTLLTLAKSDRIVGITAMAGTGLDILQKALPK 361  
QY 362 RYFPAVIAEQAATLAAGACOGAPVAIYSTFLORGQDILHDVALONLMLFALDRA 421  
DB 362 QYIDVAGIAEQAATLAAGACOGAPVAIYSTFLORGQDILHDVALONLMLFALDRA 421  
QY 422 GLVDEGPTAGAFDYSYMCIPNMLIMAPADENECROYLTTFQSHGAPSVRYPRKGP 481  
DB 422 GIVADGPTAGAFDYSYMCIPNMLIMAPADENECROYLTTFQSHGAPSVRYPRKGP 481  
QY 482 GAAT-DPTLTALTEIGKAEVRHSGRIATLAWGSAVTPAVEAGKQ-----GATVNMFPV 535  
DB 482 GVPLEPESWESLPFGKAEQLAQGDLLMLATGSKVYPMALQTAELNHEGISAIVINARFA 541  
QY 536 KFPDQALVLELARTHDVFEVTEVENYIAGAGSAINTFLQAKVLMPCNIGLPDRFVEOG 595  
DB 542 KPLDEBELIVPLARQIGKAVTPEEGCLPGFGSAIMESIQAHDLQVPLPIGVPDLVEHA 601  
QY 596 SREELSLVIGDSKGLATI-EQF 618  
DB 602 SPDESKQELGLTPROMADRILEKF 625  
RESULT 15  
OSMAYL PRELIMINARY; PRT; 722 AA.  
AC 08MAYL  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 23, Last annotation update)  
DE Deoxy-D-xylylulose-5-phosphate synthase.  
GN DXS.  
OS Morinda citrifolia (Indian mulberry).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamiales; Gentianales; Rubiaceae; Rubioideae; Morinda;  
OC Morinda.  
OK NCBI\_TaxID=43522;

RN [1]  
RP SEQUENCE FROM N.A.  
RA Han Y.S., Roytcharuk S., Verberne M.C., Heijden, Rvd, Verpoorte R.,  
RT "Cloning and characterization of a cDNA encoding 1-deoxy-D-xylylulose 5-  
RT phosphate synthase from Morinda citrifolia."  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
RM EMBL: AF43590; AA132062.1; .  
KM InterPro: IPR005477; Dxs.  
DR InterPro: IPR005476; Transketolase\_C.  
DR InterPro: IPR005475; Transketolase\_CR.  
DR InterPro: IPR005474; Transketolase\_N.  
DR Pfam: PF02780; transketolase\_C.1.  
DR Pfam: PF02779; transketolase\_Pyr.1.  
DR TIGRfam: TIGR00204; dxs.1.  
DR PROSITE: PS00801; TRANSKETOLASE\_1; 1.  
DR PROSITE: PS00802; TRANSKETOLASE\_2; 1.  
SQ SEQUENCE 722 AA; 77970 MW; 58231942EAF9C767 CRC64;  
Query Match 44.5%; Score 1430.5; DB 10; Length 722;  
Best Local Similarity 47.1%; Pred. No. 1.4e-95;  
Matches 302; Conservative 104; Mismatches 198; Indels 37; Gaps 12;  
QY 8 PLKNTHTPADIRALSK-----DQLQQLADREYRGLTHTVSISGHPAAGTVELTVA 61  
DB 72 PLDITNYPPIHMKNLSTKSHYLIQELKQALDELRSDIIFVSKTGHGSLGVELTVA 131  
QY 62 LHYVNTVPDQVWDVGHQAYPHKILTGKREMPTRITLGSVAFAPADESEYDAFGVGH 121  
DB 132 LHYAFNCPQDKIYWDVGHQAYPHKILTGKRSKQTLRTQYIAGFPDRDSVDARAGH 191  
QY 122 SSTISIAALGMAIASQSGEDKKVAIIGDGSITGMAYEAMNAGDVNANLVLINDND 181  
DB 192 SSTISIAAGLGMAYGRDLGNNNVISVIGGAVTAAQAYEAMNAGDLNANLVLINDN 251  
QY 182 MSISP-----PYGAMNNYLTQVLSKPYSSVREESKKALAKM-PSVWEAKRTEHY 231  
DB 252 QVSLPTGLDPAITPGALSSALTTLQASPKFQLEBAASITKQIGPQAEVAAKNDY 311  
QY 232 VKAMIVP--GLFPEELGPNVFGPIGDHVMVLSTLENLXDL--TGPVFLAVVTKKGY 287  
DB 312 ARGILASGTFEBELGLYITGVVDGHNIEDLITFDKATMPAPGVLIHYTEKKG 371  
QY 288 APAEKDPPLAHGVPAFPT--KQLEPKAASPHPTTYTEVFGRLCDMAADERLGITPA 345  
DB 372 PPAEAAADRNGGVKPDVAVSGQPKSK--SPILSYQFAEALIKEAELDNKIVAIHAA 428  
QY 346 MREGSGIVESQKPEPFRYPDVAIAEQAATLAAGACOGAPVAIYSTFLORGQDILH 405  
DB 429 MGGGTGLNTFQKLFPERCEPVGIAEGHAYTFAAGLITBGLKPCALYSTFLORGQD 488  
QY 406 DVALONLMLFALDRAGLVGPDPPTAGAFDYSYMCIPNMLIMAPADENECROYLTTF 465  
DB 489 DVDLQKLPVYFANDRAGLVGADGPTGCAFDVAYMCLSNIMAPADEALEHNVATAA 548  
QY 466 Q-HHGPAASVRYPRKGPAGALDPTL--TALETGKAEVRHSGRIATLAWGSAVTPAVEAG 522  
DB 549 TIDRPPCFRFPKNGKIGAKLPDKNKMLIGIKGRLITBGSVAIIGYALVQQLGAA 608  
QY 523 KQD-----GATVNMFPVFPDQALVLELARTHDVFEVTEVENYIAGAGSAINTFLQAK 577  
DB 609 EMKLNKIKPTIADAFCKRDLDELKRLAKHEHLLIYVEGSI-GSPGSHVHPLSLSG 667  
QY 578 VL--MPVCNIGLPDRFVEOGSREELSLVIGDSKGLATI 615  
DB 668 LLDGPIKRLSMVLPDRYIDGSPADQIEAGLSRHICATV 708

Search completed: January 29, 2004, 15:54:02  
Job time : 49.2142 secs



DR N-PSDB; ABK83261.

XX New high growth methanotrophic bacterial strain, useful for producing  
PT single cell proteins, grows on a C1 carbon substrate, and comprises a  
PT functional gene encoding in Embden-Meyerhof carbon pathway

PS Claim 11; Page 141-142; 157pp; English.

XX The invention relates to a high growth methanotrophic bacterial strain,  
CC which grows on a C1 carbon substrate e.g. methane and methanol, and  
CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a  
CC gene coding a pyrophosphate dependent phosphotransferase enzyme or a 16s  
CC RNA. The bacterial strain is useful for the production of single cell  
CC protein and for the biotransformation of a nitrogen-containing compound,  
CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the  
CC production of a feed product comprising a protein, carbohydrates and a  
CC pigment and for reducing oxygen demand, for removing nitrates and  
CC nitrites in methane-containing environments such as landfills, waste  
CC water treatment systems or anywhere that methane, oxygen and nitrates are  
CC present. The bacterial strain of the invention can be used as a  
CC denitrifying agent for the conversion of nitrate or nitrite to nitrous  
CC oxide with methane or methanol as a carbon source. It is also used in the  
CC production of biomass including proteins, carbohydrates and a wide  
CC variety of pigments (particularly for isoprenoid pigments for the  
CC purpose of generating animal feeds), in production of terpenoid and  
CC carotenoid compounds, useful as pigments and as monomers in polymeric  
CC materials and in production of exopolysaccharides at high levels.  
CC Sequences ABG61551-ABG61590 represent high growth methanotrophic  
CC bacterial strain proteins of the invention.

SO Sequence 394 AA;

Query Match 100.0%; Score 1985; DB 23; Length 394;

Best Local Similarity 100.0%; Pred. No. 1.8e-185;

Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MKGICILGATGSGTSTLVVARRHPKRYQVVALTANGNIDALYEQCLAHHPYAVVMES 60
DB 1 MKGICILGATGSGTSTLVVARRHPKRYQVVALTANGNIDALYEQCLAHHPYAVVMES 60
QY 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
DB 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
QY 121 VILANKKALVMSGOIFMQAVSDSGAVLPLIDSEHNALFOCPAGYTPGHTAKQARRILLT 180
DB 121 VILANKKALVMSGOIFMQAVSDSGAVLPLIDSEHNALFOCPAGYTPGHTAKQARRILLT 180
QY 181 ASGGPFRRTPIETLSSTVTPDQAVAHPKMDGKRSVDSATMMNKGLELIEACLLFNMSPD 240
DB 181 ASGGPFRRTPIETLSSTVTPDQAVAHPKMDGKRSVDSATMMNKGLELIEACLLFNMSPD 240
QY 241 QIEVVHPOSIIHSMVDYVDGSVLAQMGNDPMDRTPIAHNAWAMPERPDSGVAPLDIFEVGH 300
DB 241 QIEVVHPOSIIHSMVDYVDGSVLAQMGNDPMDRTPIAHNAWAMPERPDSGVAPLDIFEVGH 300
QY 301 MPEKEDLRFPCPLRLAYEALIKSGGIMPTVLANAEIYEAFLNEEVKFTDIAVIERISM 360
DB 301 MPEKEDLRFPCPLRLAYEALIKSGGIMPTVLANAEIYEAFLNEEVKFTDIAVIERISM 360
QY 361 AQFKPDDAGSLVLTQADODARBARVARDIIKTLVA 394
DB 361 AQFKPDDAGSLVLTQADODARBARVARDIIKTLVA 394

```

RESULT 2

AAE22302

ID AAE22302 standard; Protein, 394 AA.

AC AAE22302;

XX 25-JUL-2002 (first entry)

DT

XX

DE Methylomonas 16a sp. D-1-deoxyxyjulose-5-phosphate reductoisomerase.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;  
KM anti-oxidant; steroid; flavour; fragrance; electro-optic application;  
KM aquaculture; enzyme; D-1-deoxyxyjulose-5-phosphate reductoisomerase; Dxr.  
XX Methylomonas 16a sp.

FN W0200218617-A2.

PD 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US27420.

PF 01-SEP-2000; 2000US-229858P.

PR 01-SEP-2000; 2000US-229907P.

XX (DUPO) DU PONT DE NEMOURS & CO B. I.

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

PI Odcom JM, Plocatagis SK, Rouviere PE;

DR WPI; 2002-351711/38.

XX N-PSDB; AAD35500.

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by  
PT using microorganisms having a nucleic acid molecule encoding enzymes in  
PT the carotenoid biosynthetic pathway and which metabolize single carbon  
PT substrates

PS Claim 42; Page 112-114; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds.  
CC The method comprises a transformed metabolising host cell, comprising  
CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule  
CC encoding an enzyme in the carotenoid biosynthetic pathway, under the  
CC control of regulatory sequences, and contacting the host cell with carbon  
CC substrate to produce a carotenoid compound. The method is useful for  
CC producing carotenoid compounds such as antheraxanthin and astaxanthin, by  
CC using microorganisms having a nucleic acid molecule encoding enzymes in  
CC the carotenoid biosynthetic pathway and which metabolize single carbon  
CC substrates. The carotenoids have potent anti-oxidant properties useful in  
CC diet, and aquaculture elements. The carotenoids are also useful as  
CC intermediates in the synthesis of steroids flavours and fragrances and  
CC compounds for potential electro-optic applications. The present sequence  
CC is Methylomonas 16a sp. D-1-deoxyxyjulose-5-phosphate reductoisomerase  
CC (Dxr) enzyme used in the invention.

SO Sequence 394 AA;

Query Match 100.0%; Score 1985; DB 23; Length 394;

Best Local Similarity 100.0%; Pred. No. 1.8e-185;

Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MKGICILGATGSGTSTLVVARRHPKRYQVVALTANGNIDALYEQCLAHHPYAVVMES 60
DB 1 MKGICILGATGSGTSTLVVARRHPKRYQVVALTANGNIDALYEQCLAHHPYAVVMES 60
QY 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
DB 61 KVAEFKORIAASPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARKAKGT 120
QY 121 VILANKKALVMSGOIFMQAVSDSGAVLPLIDSEHNALFOCPAGYTPGHTAKQARRILLT 180
DB 121 VILANKKALVMSGOIFMQAVSDSGAVLPLIDSEHNALFOCPAGYTPGHTAKQARRILLT 180
QY 181 ASGGPFRRTPIETLSSTVTPDQAVAHPKMDGKRSVDSATMMNKGLELIEACLLFNMSPD 240
DB 181 ASGGPFRRTPIETLSSTVTPDQAVAHPKMDGKRSVDSATMMNKGLELIEACLLFNMSPD 240
QY 241 QIEVVHPOSIIHSMVDYVDGSVLAQMGNDPMDRTPIAHNAWAMPERPDSGVAPLDIFEVGH 300
DB 241 QIEVVHPOSIIHSMVDYVDGSVLAQMGNDPMDRTPIAHNAWAMPERPDSGVAPLDIFEVGH 300

```

QY 301 MDEKPDLRKPCICRLAYEAIAKSGIMPTVLANANEIAVEAFINBEVKETDIAVITIERSM 360  
 DB 301 MDEKPDLRKPCICRLAYEAIAKSGIMPTVLANANEIAVEAFINBEVKETDIAVITIERSM 360  
 QY 361 AQFKPDDAGSLIELVLQADQDAREVARDIIKTLVA 394  
 DB 361 AQFKPDDAGSLIELVLQADQDAREVARDIIKTLVA 394

RESULT 3  
 AAU80326  
 ID AAU80326 standard; Protein; 394 AA.  
 AC AAU80326;  
 XX 15-JUL-2002 (first entry)  
 DT  
 DE Methylomonas 16a ORF2 dxr protein sequence.  
 XX  
 XX Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
 KW keratened; pigment; flavour; fragrance; open reading frame 2; ORF2;  
 KM dxr; 1-deoxyxylulose-5-phosphate reductoisomerase enzyme.  
 OS Methylomonas sp.  
 XX  
 XX MO200220733-A2.  
 XX  
 XX 14-MAR-2002.  
 XX  
 XX 29-AUG-2001; 2001WO-US26852.  
 XX  
 XX 01-SEP-2000; 2000US-229907P.  
 XX  
 XX (DUPO ) DU PONT DE NEMOURS & CO E. I.  
 PA  
 PI Cheng Q, Koffas M, Norton KC, Odom JX, Picataggio SK, Rouviere PE;  
 PI Schenzle A, Tomb J;  
 XX WPI: 2002-383051/41.  
 DR N-PSDB; ABR50082.  
 XX  
 PT Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,  
 PT isolated from Methylomonas 16a, useful for the production of isoprenoid  
 PT compounds -  
 PS  
 PS Claim 4; Page 68-70; 84pp; English.  
 XX  
 CC The present invention relates to a new nucleic acid molecule encoding  
 CC an isoprenoid biosynthetic enzyme isolated from Methylomonas 16a.  
 CC The invention is useful for obtaining a nucleic acid molecule  
 CC encoding an isoprenoid compound biosynthetic enzyme, and for the  
 CC microbial production of isoprenoid compounds. The molecules of the  
 CC invention are also useful for regulating isoprenoid biosynthesis in an  
 CC organism and for producing recombinant organisms for producing various  
 CC isoprenoid compounds. The nucleic acid is also useful for feed additive,  
 CC for the production of kerateneds and their derivatives, isoprenoid  
 CC intermediates, and as pure products useful as pigments, flavour and  
 CC fragrances. The present amino acid sequence represents the Methylomonas  
 CC 16a open reading frame 2 (ORF2) dxr (1-deoxyxylulose-5-phosphate  
 CC reductoisomerase enzyme) protein of the invention, as described above.  
 XX  
 XX Sequence 394 AA;

Query Match 100.0%; Score 1985; DB 23; Length 394;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-185;  
 Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGICITGATGSGTGTATVVAHPKQYVATLTANGNDALYECIAHHPYAAVWMS 60  
 DB 1 MGICITGATGSGTGTATVVAHPKQYVATLTANGNDALYECIAHHPYAAVWMS 60  
 QY 61 KVAEFKQRIAAAPVADIKVLISGSEALQGVATLENTVTAALVGAALPTLAAAKAGKT 120

DB 61 KVAEFKQRIAAAPVADIKVLISGSEALQGVATLENTVTAALVGAALPTLAAAKAGKT 120  
 QY 121 VLANKKALVWVGQIFMQAVSDGAVLLPIDSENAIIFQCPAGYTPGHTAKQARRILT 180  
 DB 121 VLANKKALVWVGQIFMQAVSDGAVLLPIDSENAIIFQCPAGYTPGHTAKQARRILT 180  
 QY 181 ASGPFRTPTPIETLSVTPDQAVAHPRMDWKRKISVSATMNGKLEIECLTFNNEPD 240  
 DB 181 ASGPFRTPTPIETLSVTPDQAVAHPRMDWKRKISVSATMNGKLEIECLTFNNEPD 240  
 QY 241 QIEVVIHPQSIITHSMVDYVDSVLAQMGNDPMDRTPIAHAMAPRFDGVAFLDIFEVGH 300  
 DB 241 QIEVVIHPQSIITHSMVDYVDSVLAQMGNDPMDRTPIAHAMAPRFDGVAFLDIFEVGH 300  
 QY 301 MDEKPDLRKPCICRLAYEAIAKSGIMPTVLANANEIAVEAFINBEVKETDIAVITIERSM 360  
 DB 301 MDEKPDLRKPCICRLAYEAIAKSGIMPTVLANANEIAVEAFINBEVKETDIAVITIERSM 360  
 QY 361 AQFKPDDAGSLIELVLQADQDAREVARDIIKTLVA 394  
 DB 361 AQFKPDDAGSLIELVLQADQDAREVARDIIKTLVA 394

RESULT 4  
 AA021884  
 ID AA021884 standard; Protein; 396 AA.  
 AC AA021884;  
 XX  
 XX 13-SEP-2002 (first entry)  
 DT  
 DE Isoprenoid related protein sequence SEQ ID No 117.  
 XX  
 XX Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
 KM decaprenyl diphosphate synthase.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 XX MO200226933-A2.  
 XX  
 XX 04-APR-2002.  
 XX  
 XX 28-SEP-2001; 2001WO-US30328.  
 XX  
 XX 29-SEP-2000; 2000US-236580P.  
 XX  
 XX (CRGI ) CARGILL INC.  
 PA  
 PI Gokarn R, Jessen H, Zidwick MJ;  
 PI WPI: 2002-416480/44.  
 DR  
 PT Substantially pure polypeptides having e.g. 1-deoxyxylulose-5-phosphate synthase activity,  
 PT production of isoprenoids, especially CoQ(10), useful for the  
 PT production of isoprenoids, especially CoQ(10).  
 PS  
 PS Disclosure; Fig 32; 246pp; English.  
 XX  
 CC The invention relates to methods and materials for the production of  
 CC isoprenoids. More particularly the invention provides isolated nucleic  
 CC acids, substantially pure polypeptides, host cells, and methods for  
 CC producing various isoprenoid compounds. The polypeptides are useful for  
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure  
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
 CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
 CC increasing production of CoQ(10) in a cell having endogenous DDS  
 CC activity. This sequence represents a protein relating to the isoprenoid  
 CC production of the invention.  
 XX  
 XX Sequence 396 AA;

Query Match 61.2%; Score 1215.5; DB 23; Length 396;

Best Local Similarity 64.1%; Pred. No. 4,1e-110; Matches 245; Conservative 48; Mismatches 88; Indels 1; Gaps 1;

4 ICLGATGSGVSTLDVVAARHPDKYQVVALFANGNIDALYEQCLAHPEYAAVVMESKVA 63  
 7 ISVLTGTSIGSTLDVVRHPRDRTEAPALTPSKRLAEALCLHPRPYAAVPEQAAAI 66

64 EFKQRIASPVDIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAAAKGKTVLL 123  
 67 ALQGSILAAAGIR-TRVLFGEQALCEVASAPEVDMAAIVGAAGLPSTLAAVEAGKRVLL 125

124 ANKEALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCMRPGYTPGHAKQARILLTASG 183  
 126 ANKEALVMSGALFMQAVKSGAVLLPIDSEHNAIFQSLPRNADGLERGVARRILLTASG 185

184 GPFRRPIETLSVTPDQAVAHPKMDMGRKISVDSATMNMKGLELIEACLLFNNEPDQIE 243  
 186 GPFRRPIETLSVTPDQAVAHPKMDMGRKISVDSATMNMKGLELIEACLLFNNEPDQIE 245

244 VTIHQSTIHSKVDYVDSVLAQMGNDPRTPIAHAMAMBERPDSGVAPLDIFEVGMDP 303  
 246 VTIHQSTIHSKVDYVDSVLAQMGNDPRTPIAHAMAMBERPDSGVAPLDIFEVGMDP 305

304 EKPDLKRFPCILAEYAIKSGIMPTVLANAEIAVEAFINEEVFTDIAVIERMAQF 363  
 306 QRPDQRFPCILAEYAIKSGIMPTVLANAEIAVEAFINEEVFTDIAVIERMAQF 365

364 KPDDAGSLVLAQDADAREVA 385  
 366 AVTAVESLDQVLAADRRASVA 387

RESULT 5  
 AAO21886  
 ID AAO21886 standard; Protein; 394 AA.

AAO21886;  
 13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 119.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
 decaprenyl diphosphate synthase.

Neisseria meningitidis.  
 WO200226933-A2.

04-APR-2002.  
 28-SEP-2001; 2001WO-US30328.  
 29-SEP-2000; 2000US-236580P.

(CRGI ) CARGILL INC.  
 Gokarn R, Jessen H, Zidwick MJ;  
 WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,  
 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
 production of isoprenoids, especially CoQ(10)

Disclosure; Fig 32; 246pp; English.

The invention relates to methods and materials for the production of  
 isoprenoids. More particularly the invention provides isolated nucleic  
 acids, substantially pure polypeptides, host cells, and methods for  
 producing various isoprenoid compounds. The polypeptides are useful for  
 the production of isoprenoids, especially CoQ(10). Expressing the pure  
 polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)

CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
 CC increasing production of CoQ(10) in a cell having endogenous DDS  
 CC activity. This sequence represents a protein relating to the isoprenoid  
 CC production of the invention.

XX Sequence 394 AA;

Query Match 53.0%; Score 1052.5; DB 23; Length 394;  
 Best Local Similarity 55.5%; Pred. No. 3.7e-94;  
 Matches 216; Conservative 49; Mismatches 123; Indels 1; Gaps 1;

4 ICLGATGSGVSTLDVVAARHPDKYQVVALFANGNIDALYEQCLAHPEYAAVVMESKVA 63  
 6 LTITGTSIGSESTLDVVSHPERFVPLAGHQVKLAQCTFPERYAVVADAHAA 65

64 EFKQRIASPVDIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAAAKGKTVLL 123  
 66 RLBALIKRDGTA-QVTLHGAQALVDVASADEVSGMCAIVGAVLPALAAQKQKTIYL 124

124 ANKEALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCMRPGYTPGHAKQARILLTASG 183  
 125 ANKETLVVSGALFMEETARANGAAVLPVDSHNAVQVLPEDYGRLEHNGIASIILLTASG 184

184 GPFRRPIETLSVTPDQAVAHPKMDMGRKISVDSATMNMKGLELIEACLLFNNEPDQIE 243  
 185 GPFRLADLNTFDSITPDQAVAHPKMDMGRKISVDSATMNMKGLELIEAHWLFNCPDYLE 244

244 VTIHQSTIHSKVDYVDSVLAQMGNDPRTPIAHAMAMBERPDSGVAPLDIFEVGMDP 303  
 245 VTIHQSTIHSKVDYVDSVLAQMGNDPRTPIAHAMAMBERPDSGVAPLDIFEVGMDP 304

304 EKPDLKRFPCILAEYAIKSGIMPTVLANAEIAVEAFINEEVFTDIAVIERMAQF 363  
 305 QRPDQRFPCILAEYAIKSGIMPTVLANAEIAVEAFINEEVFTDIAVIERMAQF 364

364 KPDDAGSLVLAQDADAREVA 392  
 365 FSDGIDGIGLLAODARTRAQARAFTIGTL 393

RESULT 6  
 AAO21878  
 ID AAO21878 standard; Protein; 398 AA.

AAO21878;  
 13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 100.

Isoprenoid related protein sequence SEQ ID No 100.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
 decaprenyl diphosphate synthase.

Escherichia coli.  
 WO200226933-A2.

04-APR-2002.  
 28-SEP-2001; 2001WO-US30328.  
 29-SEP-2000; 2000US-236580P.

(CRGI ) CARGILL INC.  
 Gokarn R, Jessen H, Zidwick MJ;  
 WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,  
 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
 production of isoprenoids, especially CoQ(10)

PS Disclosure; Fig 27; 246pp; English.

The invention relates to methods and materials for the production of isoprenoids. More particularly the invention provides isolated nucleic acids, substantially pure polypeptides, host cells, and methods for producing various isoprenoid compounds. The polypeptides are useful for the production of isoprenoids, especially CoQ(10). Expressing the pure polypeptides, which has 1-deoxyxylyulose-5-phosphate synthase (DXS) activity or decaprenyl diphosphate synthase (DDS) activity, is useful for increasing production of CoQ(10) in a cell having endogenous DDS activity. This sequence represents a protein relating to the isoprenoid production of the invention.

Sequence 398 AA;

Query Match 52.6%; Score 1045; DB 23; Length 398;  
Best Local Similarity 54.9%; Pred. No. 2e-93;  
Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3

YY MKGICITLGTSGISGVSTLDVVAHFHDKTQVVALTANGNIDALYEQLAAHPYAVVMES 60  
DB 1 MKGLTLTGSTSIGSTLDVVRHNHEHRVVALVAGKIVTRVBQCLESPPYAVMDRA 60  
YY 61 KVAEFKQRIASPVADIVYLSSGSALKQCVATLENDVTMAAIVGAAGLIPTAAAKAGT 120  
DB 61 S-AKLTKMTLDOOGSRTEVLSSQQAACDMAALEDDVQMAAIVGAAGLIPTAAIRAGKT 119  
YY 121 VLLANKELAWSGOIFMOAVSPSGAVLLPIDSEHNAIFOCM--PAGYTPGHAKQAR--- 175  
DB 120 ILLANKSELVTCGRIFMDVAVKOSRKQQLFPDSBHALLQSLSQPIGHNLGYDLEONGYV 179  
YY 176 RILLTASGGPRFRPEIETLSVTPDQAAVHAHPKDMGRKISVDSATWNGKELEIACILF 235  
DB 180 SILTLTSGGPFRRETPRLDLATWTLPQACRHPWSMGKRKISVDSATPMNKGLEIARWL 239  
YY 236 NMEPDQIVVVIHQSIHTSMVDYDVGSTLAQGNPDMTPIINAAWAEERPLSGAAPDI 295  
DB 240 NASASQKEVLLIHPOSVIHSWRVQGSVLAOLGSDPKETPIAHYAAMPNRVNSGVKPLDF 299  
YY 296 FEVGHMDEEKEDLKRFPCPLRAYEAIKSGGIIMPVTLANAEIAVEAFINEVEKFTDIAYI 355  
DB 300 CKLSALTFAAPDYDRYPCFLKAMEAFBEGQAATTNANNEITVAAPFLAQQRIFTDIAL 359  
YY 356 IERSMAPKXPDDAGSLFVLAQQDAFEVARDIIKTLVA 394  
DB 360 NLSTLERKMDEREPQCVDVLTSYDANAREVARREVETLAS 398

RESULT 7  
AA021885  
ID AA021885 standard; Protein; 398 AA.  
XX  
XX AA021885;  
XX  
XX DT 13-SEP-2002 (first entry)  
XX DE Isoprenoid related protein sequence SEQ ID NO 118.  
XX KW Isoprenoid; CoQ(10); 1-deoxyxylyulose-5-phosphate synthase; DXS; DDS;  
XX RV decaprenyl diphosphate synthase.  
XX OS Escherichia coli.  
XX PN WO200226933-A2.  
XX PD 04-APR-2002.  
XX PF 28-SEP-2001; 2001WO-US30326.  
XX PR 29-SEP-2000; 2000US-236580P.  
XX RA (CRGI ) CARGILL INC.  
XX

```

XX 12-FEB-2002; 2002MO-IB02069.
PF 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
PA Fontana MR, Pizza M, Massignani V, Monaci E;
XX WPI: 2003-058415/05.
DR N-PSDB; AB240669.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection
XX
XX Disclosure; Page 621; 815bp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP6736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
XX Sequence 411 AA;
SQ
Query Match 52.6%; Score 1043.5; DB 24; Length 411;
Best Local Similarity 55.5%; Pred. No. 3e-93;
Matches 216; Conservative 48; Mismatches 124; Indels 1; Gaps 1;
XX
QY 4 ICILGATSGISGTVLDVVAHHPDKYQVVALTANGNIDALVEOCLAHPEYAVVMESKVA 63
DB 23 LTIIGSTGSGESTLDVVAHHPDKYQVVALTANGNIDALVEOCLAHPEYAVVMESKVA 82
QY 64 EFKQIASSPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARAGKTVLL 123
DB 83 RLEALIKRDGTA-TQVHGAQALVDVASADEVSGVMCAIVGAAGLLPTLAARAGKTVLL 141
QY 124 ANKEALVMSGOIFMOWAVSDGAVLLPIDSEHNAIFQCPAGYTPGHTAKQARRILLTASG 183
DB 142 ANKETLVVSGALFMETARANGAVALPVDSHNAIFQVLPDYTDRLNEHGIDSIILLTASG 201
QY 184 GPFRRPIETLSVTPDOAVAHPRKMDGKISVDSATMANKGLIELIACLLFNNPDQIE 243
DB 202 GPFLLTDLSTFDSTIPBQAVKHPMRMGKISVDSATMANKGLIELIACLLFNNPDQIE 261
QY 244 VVIHQSIHSMVDYDVGSLAQMGNPDKRTPIAHAMAPRERFDSGVALPDIPEVGMDF 303
DB 262 VVIHQSVIHSWVRVSDGSLAQLGNPDKRTPIAHAMAPRERFDSGVALPDIPEVGMDF 321
QY 304 EKPDLEKFPCLRLAYEAIKSGIMPTVLANAEIAYEAFINEVEKFTDIIVIIERSNAQF 363
DB 322 QKPDGRFPCLKFAVEITINAGAAPCVLANAEIAYEAFINEVEKFTDIIVIIERSNAQF 381
QY 364 KPDAGSLLEVLAQADQAREVARDIITKL 392
DB 382 FSNMGDIEGLAQAQDARTRAQARAFIGTL 410
XX
RESULT 9
ABP80486
ID ABP80486 standard; Protein; 411 AA.
XX
AC ABP80486;
XX
DT 07-MAR-2003 (first entry)
XX
DB N. gonorrhoeae amino acid sequence SEQ ID 7502.
XX
KW Antibacterial; infection; vaccine; gene therapy.
XX

```

```

OS Neisseria gonorrhoeae.
XX
XX WO200279243-A2.
XX
XX 10-OCT-2002.
XX
XX 12-FEB-2002; 2002MO-IB02069.
XX
XX 12-FEB-2001; 2001GB-0003424.
XX (CHIR-) CHIRON SPA.
XX
XX Fontana MR, Pizza M, Massignani V, Monaci E;
XX WPI: 2003-058415/05.
DR N-PSDB; AB241456.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection
XX
XX Disclosure; Page 734; 815bp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP6736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
XX Sequence 411 AA;
SQ
Query Match 52.6%; Score 1043.5; DB 24; Length 411;
Best Local Similarity 55.5%; Pred. No. 3e-93;
Matches 216; Conservative 48; Mismatches 124; Indels 1; Gaps 1;
XX
QY 4 ICILGATSGISGTVLDVVAHHPDKYQVVALTANGNIDALVEOCLAHPEYAVVMESKVA 63
DB 23 LTIIGSTGSGESTLDVVAHHPDKYQVVALTANGNIDALVEOCLAHPEYAVVMESKVA 82
QY 64 EFKQIASSPVADIKVLSGSEALQVATLENDVTMAAIVGAAGLLPTLAARAGKTVLL 123
DB 83 RLEALIKRDGTA-TQVHGAQALVDVASADEVSGVMCAIVGAAGLLPTLAARAGKTVLL 141
QY 124 ANKEALVMSGOIFMOWAVSDGAVLLPIDSEHNAIFQCPAGYTPGHTAKQARRILLTASG 183
DB 142 ANKETLVVSGALFMETARANGAVALPVDSHNAIFQVLPDYTDRLNEHGIDSIILLTASG 201
QY 184 GPFRRPIETLSVTPDOAVAHPRKMDGKISVDSATMANKGLIELIACLLFNNPDQIE 243
DB 202 GPFLLTDLSTFDSTIPBQAVKHPMRMGKISVDSATMANKGLIELIACLLFNNPDQIE 261
QY 244 VVIHQSIHSMVDYDVGSLAQMGNPDKRTPIAHAMAPRERFDSGVALPDIPEVGMDF 303
DB 262 VVIHQSVIHSWVRVSDGSLAQLGNPDKRTPIAHAMAPRERFDSGVALPDIPEVGMDF 321
QY 304 EKPDLEKFPCLRLAYEAIKSGIMPTVLANAEIAYEAFINEVEKFTDIIVIIERSNAQF 363
DB 322 QKPDGRFPCLKFAVEITINAGAAPCVLANAEIAYEAFINEVEKFTDIIVIIERSNAQF 381
QY 364 KPDAGSLLEVLAQADQAREVARDIITKL 392
DB 382 FSNMGDIEGLAQAQDARTRAQARAFIGTL 410
XX
RESULT 10
AAO21889
ID AAO21889 standard; Protein; 400 AA.
XX
AC AAO21889;
XX
DT 13-SEP-2002 (first entry)
XX

```



XX Isoprenoid related protein sequence SEQ ID No 122.  
 DE Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
 XX decaprenyl diphosphate synthase.  
 KW Pasteurella multocida.  
 XX WO200226933-A2.  
 XX 04-APR-2002.  
 XX 28-SEP-2001; 2001WO-US30328.  
 XX 29-SEP-2000; 2000US-236580P.  
 XX (CRGI ) CARGILL INC.  
 XX Gokarn R, Jeesen H, Zidwick MJ;  
 XX WPI; 2002-416480/44.  
 XX Substantially pure polypeptides having e.g.,  
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
 PT production of isoprenoids, especially CoQ(10).  
 XX Disclosure; Fig 32; 246pp; English.  
 XX The invention relates to methods and materials for the production of  
 CC isoprenoids. More particularly the invention provides isolated nucleic  
 CC acids, substantially pure polypeptides, host cells, and methods for  
 CC producing various isoprenoid compounds. The polypeptides are useful for  
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure  
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
 CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
 CC increasing production of CoQ(10) in a cell having endogenous DDS  
 CC activity. This sequence represents a protein relating to the isoprenoid  
 CC production of the invention.  
 XX Sequence 400 AA;  
 SQ  
 Query Match 52.5%; Score 1042; DB 23; Length 400;  
 Best Local Similarity 53.2%; Pred. No. 4,1e-93;  
 Matches 210; Conservative 68; Mismatches 109; Indels 8; Gaps 3;  
 QY 1 MGICLTGATGSGTLDVVARHPDKQVVALTANGNDALYECGLAHPEYAVVWES 60  
 DB 2 MKKIVILSTGISTGTLSTVITHNPKIVFALVGGRNVELMFQCLTFQPSFALDDV 61  
 QY 61 KVAEFKORIAAPVADIKVLSGSEALQVATLENVDYMAALVGAAGLLPTLAAGKT 120  
 DB 62 AKNKAELKKAHQ-SGTVLAGQAICELAAHPADMMVAALVGAAGLLPTLSAVKGR 120  
 QY 121 VLLANKKALVMSGOIFMAVSDSGAVLPIIDSEHNAIFQCM-----AGTTPGHTAKQA 174  
 DB 121 VLLANKKALVLTGQGLFIDAVRESQALPVDSEHNAIFQSLPPEARQGTGCP-ISELGI 179  
 QY 175 RRIILTAGSGPRFRPIETLSSVTPOAVAHKPMGGRKISVDSATPMNKGLEIEACILF 234  
 DB 180 SKIVLTGSGPRFRPIETLSSVTPOAVAHKPMGGRKISVDSATPMNKGLEIEARML 239  
 QY 235 FNNKPDQIEVVIHPSIITHSWDYVDGSLAQMGNPDKMTPIAHAMAFERDSCVAPLD 294  
 DB 240 FNAAEEMEVIIHPSIITHSMRYIDGSVIAQMGNPDKMTPIAHMAVPSRTVAGVEPLD 299  
 QY 295 IEEVGNDEKPDLPKPFCLRLAVEAIKSGIMPTVLAANAIYEAFAINBEVKFTDIAY 354  
 DB 300 FQGLNLTLPEDYQYRPLCKLAIDFSGVAYATTAANAAEIAVASFIDMKIKFTDIAR 359  
 QY 355 IIRSNAPKPPDAGSLLELVLAQADQAREVARDII 389  
 DB 360 LMQLVVSKLPQFKIHCTIDVLVEYDKKARELSQSI 394

RESULT 11  
 ID AAO21877  
 AC AAO21877 standard; Protein; 397 AA.  
 XX AAO21877;  
 XX 13-SEP-2002 (first entry)  
 XX Isoprenoid related protein sequence SEQ ID No 99.  
 DE Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
 KW decaprenyl diphosphate synthase.  
 XX Haemophilus influenza.  
 XX WO200226933-A2.  
 XX 04-APR-2002.  
 XX 28-SEP-2001; 2001WO-US30328.  
 XX 29-SEP-2000; 2000US-236580P.  
 XX (CRGI ) CARGILL INC.  
 XX Gokarn R, Jeesen H, Zidwick MJ;  
 XX WPI; 2002-416480/44.  
 XX Substantially pure polypeptides having e.g.,  
 PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
 PT production of isoprenoids, especially CoQ(10).  
 XX Disclosure; Fig 27; 246pp; English.  
 XX The invention relates to methods and materials for the production of  
 CC isoprenoids. More particularly the invention provides isolated nucleic  
 CC acids, substantially pure polypeptides, host cells, and methods for  
 CC producing various isoprenoid compounds. The polypeptides are useful for  
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure  
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
 CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
 CC increasing production of CoQ(10) in a cell having endogenous DDS  
 CC activity. This sequence represents a protein relating to the isoprenoid  
 CC production of the invention.  
 XX Sequence 397 AA;  
 SQ  
 Query Match 50.7%; Score 1007; DB 23; Length 397;  
 Best Local Similarity 50.6%; Pred. No. 1,1e-89;  
 Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;  
 QY 2 KGICLTGATGSGTLDVVARHPDKQVVALTANGNDALYECGLAHPEYAVVWESX 61  
 DB 4 QNIVILSTGISTGTLSTVITHNPKIVFALVGGRNVELMFQCLTFQPSFALDDVNA 63  
 QY 62 VAEKFORIAAPVADIKVLSGSEALQVATLENVDYMAALVGAAGLLPTLAAGKT 121  
 DB 64 AKIIRKLIHNT-PTETLARRAICELAAHPADQIMASIVGAAGLLPTLSAVKGRV 122  
 QY 122 LLANKKALVMSGOIFMAVSDSGAVLPIIDSEHNAIFQCM-----AGTTPGHTAKQA 175  
 DB 123 LLANKKALVLTGQGLFIDAVKNGSKLPVDSEHNAIFQSLPPEARQGTGCP-ISELGV 181  
 QY 176 RRIILTAGSGPRFRPIETLSSVTPOAVAHKPMGGRKISVDSATPMNKGLEIEACILF 235  
 DB 182 KIILTAGSGPRFRPIETLSSVTPOAVAHKPMGGRKISVDSATPMNKGLEIEARML 241  
 QY 236 FNNKPDQIEVVIHPSIITHSWDYVDGSLAQMGNPDKMTPIAHAMAFERDSCVAPLDI 295  
 DB 242 NAAEEMEVIIHPSIITHSMRYIDGSVIAQMGNPDKMTPIAHMAVPSRTVAGVEPLD 301

QY 296 FEVGHMDPEKPKRPPCLRLAYEALIKSGIMPTVLNANEIAYEALINEEYKTDIAVI 355  
 DB 302 FKIKELTFIEPDPFNRYPLKLAIDAFAGAYATTANANAEIAYAVAFDRQIGFMDIAKI 361  
 QY 356 IERSMAQFKPPDAGSLLEVLQADQDAREVARDIIR 390  
 DB 362 NSKTERISPTTIOINIDVLEIDQAREIAKTIILR 396

## RESULT 12

AAO21887  
 ID AAO21887 standard; Protein; 397 AA.

AAO21887;

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SHQ ID No 120.

Isoprenoid, CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;  
 decaprenyl diphosphate synthase.

Haemophilus influenza.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001MO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI ) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,  
 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
 production of isoprenoids, especially CoQ(10)

Disclosure: Fig 32, 246pp; English.

The invention relates to methods and materials for the production of  
 isoprenoids. More particularly the invention provides isolated nucleic  
 acids, substantially pure polypeptides, host cells, and methods for  
 producing various isoprenoid compounds. The polypeptides are useful for  
 the production of isoprenoids, especially CoQ(10). Expressing the pure  
 polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
 activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
 increasing production of CoQ(10) in a cell having endogenous DDS  
 activity. This sequence represents a protein relating to the isoprenoid  
 production of the invention.

Sequence 397 AA;

Query Match 50.7%; Score 1007; DB 23; Length 397;

Best Local Similarity 50.6%; Pred. No. 1,1e-89;

Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;

2 KGICILCATSGISGVTLVDVAVARHPDKYQVALTANKNIDLYRQCLAHNEEYAVVMESEK 61

4 QNTVILGISTOSIGKSTLSVENNPQKXHAFLVGGKVAEMFDCIKRPHFALDDVNA 63

62 VABPKORIASPVADIVLGSSEALQOVATLENDVTMAIVAGLPTLAATAKAGKY 121

64 AKILREGLIAHRTPTREVLAGRRAICGLAHPADQIMATIVAGLPTLSAVKAGKY 122

122 LLANKEALVMSGQIFKQAVSDSGAVILPIDSSENAITQCM-----AGTTPGHTAKQAR 175

123 LLANKESLIVCGQFLIDAVKYGSKLIPVDSSENAITQSLPBAQKXIGFCP-LSEIGVS 181

QY 176 RILITASGPPRRPTPTITLSTVPDQAVARPKMDMKGISVDSATNMKGLIELIAQLLF 235  
 DB 182 KIILTGSGGPRRYPLPLPQFNITPEQAVAHNMGKGISVDSATNMKGLIEYEAHMLF 241  
 QY 236 NMEPDQIEVVIHPSIISHMWYDVYDGSVLQAGMPDKRTPIAHAMAPBEPDSGVAPLDI 295  
 DB 242 NASAEHEVVIHPSIISHMWYDVYDGSVLQAGMPDKRTPIAHAMAPBEPDSGVAPLDI 301  
 QY 296 FEVGHMDPEKPKRPPCLRLAYEALIKSGIMPTVLNANEIAYEALINEEYKTDIAVI 355  
 DB 302 FKIKELTFIEPDPFNRYPLKLAIDAFAGAYATTANANAEIAYAVAFDRQIGFMDIAKI 361  
 QY 356 IERSMAQFKPPDAGSLLEVLQADQDAREVARDIIR 390  
 DB 362 NSKTERISPTTIOINIDVLEIDQAREIAKTIILR 396

## RESULT 13

ABB09436  
 ID ABB09436 standard; Protein; 397 AA.

ABB09436;

01-JUL-2002 (first entry)

H. influenzae DXR reductoisomerase enzyme polypeptide sequence.

DXR, reductoisomerase; enzyme; non-mevalonate isoprenoid;

meningitis; pneumonia; conjunctivitis; bacteremia; sinusitis;

pleural empyema; endocarditis; epiglottitis.

Haemophilus influenzae.

MO200211673-A2.

14-FEB-2002.

09-AUG-2001; 2001MO-US24950.

09-AUG-2000; 2000US-223909P.

(SMIR ) SMITHKLINE BEECHAM CORP.

(SMIR ) SMITHKLINE BEECHAM PLC.

Jaworski DD, Payne DJ, Slater-Radostki CE, Yan K;

WPI; 2002-241698/29.

Modulating Haemophilus influenzae DXR reductoisomerase enzyme activity,  
 useful for treating mammals or tissues infected with H. influenzae  
 (e.g. ear infections or pneumonia) by contacting the enzyme with a  
 modulator of its activity

Claim 6(i); Page 5; 44pp; English.

The invention relates to modulating an activity of a DXR reductoisomerase  
 enzyme of Haemophilus influenzae, comprising contacting the enzyme with a  
 compound that modulates non-mevalonate isoprenoid biosynthesis -  
 synthesis of menaquinone or ubiquinone. Compounds of the invention act as  
 vincloides. The method is useful for treating a mammal or mammalian tissue  
 infected with H. influenzae having DXR reductoisomerase enzyme, e.g. a  
 human or a domestic animal. In particular, the method is useful for  
 treating ear infections, conjunctivitis, meningitis, pneumonia,  
 conjunctivitis, bacteremia, sinusitis, pleural empyema, endocarditis and  
 epiglottitis. The current sequence represents the H. influenzae DXR  
 reductoisomerase enzyme polypeptide sequence.

Note: in the sequence listing, on page 40-44 of the specification, there  
 is given an amino acid sequence that is also described as SEQ ID 2, which  
 is the identifier of the current sequence. This sequence is given in  
 record ABB09437, and contains the amino acids given in the current  
 sequence but with a large insertion of extra amino acids that are not

CC encoded in the polynucleotide that encodes the enzyme of the invention  
 CC (see AB40400).

XX Sequence 397 AA;

Query Match 50.7%; Score 1007; DB 23; Length 397;

Best Local Similarity 50.6%; Pred. No. 1.1e-89; Indels 8; Gaps 3;

Matches 200; Conservative 76; Mismatches 111;

2 KQICLVATGSGIGVSTLDVYARHPDKYQVVALTANGNIDALYEQCLAHHPYAVVWMSK 61

4 QNIVILGSGISIGKSTLSYIENNPKYVHAFALVGKNNVAMFQCKRPHPAALDDVNA 63

62 VAFKRIASPVADIKVLSGSEALQOVATLENVDVMAIVGAGLPTLAAAKAGTV 121

64 AKILREKLAJAHNI-PTVELAGRAICELAAHPADIDMASIVGAGLPTLSAVVAGKRV 122

122 LLAANKALVMSGQIFMGAVSDSGAVLLPTDSEHNAIFQCP-----AGYTPGHTAKOAR 175

123 LLAANKSLVTCGLPTDAVAVNGSKLLPVDSEHNAIFQSLPPEAQEKIGFCP-LSHIGVS 181

176 RILFNASGSPFRPTPIETLSVTPDQAVAHAPKMDGRKISVDSATMNGLELIEACLLF 235

182 KIILTGSGGSPFRPTPIETLSVTPDQAVAHAPKMDGRKISVDSATMNGLELIEACLLF 241

236 NNEPDQIEVVIHPOSIHSMNDVYDGSVLAQKGNPDKRTPIAHAMAPRPSGVAPLDI 295

242 NASAEVMEVLIHQSIHSMNRKIDVDSVLTQGNPDKRTPIAHAMAPRPSGVAPLDI 301

296 FEVGHNDPEKPKDLKRPCKRLAVEAIKSGIMPTVLAANEIAVEAFINBEVFTDIATV 355

302 FIKELFTLEPDRNRPNLKLAIDAFAGVATTANNAEIAVAFLDRQIGFDIATV 361

356 IERSMAQFEPDAGSLELVQADQAREVARDIITK 390

362 NSKTERISPYTIQIDVLEIDAQREIAKTLLR 396

RESULT 14

AAO21879

AAO21879 standard; Protein; 388 AA.

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 101.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;

decaprenyl diphosphate synthase.

Zymonas mobilis.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001MO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

WPI; 2002-416480/44.

Substantially pure polypeptides having e.g.,

1-deoxyxylulose-5-phosphate synthase activity, useful for the

production of isoprenoids, especially CoQ(10)

Disclosure; Fig 27; 246pp; English.

CC The invention relates to methods and materials for the production of  
 CC isoprenoids. More particularly the invention provides isolated nucleic  
 CC acids, substantially pure polypeptides, host cells, and methods for  
 CC producing various isoprenoid compounds. The polypeptides are useful for  
 CC the production of isoprenoids, especially CoQ(10). Expressing the pure  
 CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
 CC activity or decaprenyl diphosphate synthase (DDS) activity, is useful for  
 CC increasing production of CoQ(10) in a cell having endogenous DDS  
 CC activity. This sequence represents a protein relating to the isoprenoid  
 CC production of the invention.

XX Sequence 388 AA;

Query Match 50.4%; Score 999.5; DB 23; Length 388;

Best Local Similarity 50.9%; Pred. No. 5.6e-89; Indels 9; Gaps 3;

Matches 199; Conservative 74; Mismatches 109;

4 ICILGATGSGIGVSTLDVYARHPDKYQVVALTANGNIDALYEQCLAHHPYAVVWMSKVA 63

7 VYVIGATGSGIGVSTLDVYARHPDKYQVVALTANGNIDALYEQCLAHHPYAVVWMSKVA 66

64 BFKORIASPVADIKVLSGSEALQOVATLENVDVMAIVGAGLPTLAAAKAGTV 123

67 DLKELTAGS---SVAAAGADALVEAAMW-CADWTMAIIGCAGIKATLAIIRKGTVAL 122

124 ANKEALVMSGQIFMGAVSDSGAVLLPTDSEHNAIFQCPAPYTPGHTAKOARRIILITAGS 183

123 ANKESLVASGGLMDDAVNEHGTLLPEVDSHNAIFQCP-----PHNNDYVARIILITAGS 177

184 GPFRTPIETLSVTPDQAVAHAPKMDGRKISVDSATMNGLELIEACLLFENNEPDQIE 243

178 GPFRTPIETLSVTPDQAVAHAPKMDGRKISVDSATMNGLELIEACLLFENNEPDQIE 237

244 VVIHQSIHSMNDVYDGSVLAQKGNPDKRTPIAHAMAPRPSGVAPLDIPEVGHNDP 303

238 ILVHQSIHSMNDVYDGSVLAQKGNPDKRTPIAHAMAPRPSGVAPLDIPEVGHNDP 297

304 EKPDLKRPCKRLAVEAIKSGIMPTVLAANEIAVEAFINBEVFTDIATVIERMAQF 363

298 EAPDYERFPALTIMESIKSGARPAVANNANETAVAFDLKIKIGFDIATVETKIDHY 357

364 KPDAGSLELVQADQAREVARDIITKVA 394

358 TPATPSSLEVDVFAIDNEARIQAAALMESLPA 388

RESULT 15

AAO21883

AAO21883 standard; Protein; 388 AA.

13-SEP-2002 (first entry)

Isoprenoid related protein sequence SEQ ID No 116.

Isoprenoid; CoQ(10); 1-deoxyxylulose-5-phosphate synthase; DXS; DDS;

decaprenyl diphosphate synthase.

Zymonas mobilis.

MO200226933-A2.

04-APR-2002.

28-SEP-2001; 2001MO-US30328.

29-SEP-2000; 2000US-236580P.

(CRGI) CARGILL INC.

Gokarn R, Jessen H, Zidwick MJ;

DR MPI; 2002-416480/44.

XX Substantially pure polypeptides having e.g.,  
PT 1-deoxyxylulose-5-phosphate synthase activity, useful for the  
PT production of isoprenoids, especially CoQ(10).  
XX

PS Disclosure; Fig 32; 246pp; English.

XX  
CC The invention relates to methods and materials for the production of  
CC isoprenoids. More particularly the invention provides isolated nucleic  
CC acids, substantially pure polypeptides, host cells, and methods for  
CC producing various isoprenoid compounds. The polypeptides are useful for  
CC the production of isoprenoids, especially CoQ(10). Expressing the pure  
CC polypeptides, which has 1-deoxyxylulose-5-phosphate synthase (DXS)  
CC activity or decaprenyl diphosphate synthase (DPS) activity, is useful for  
CC increasing production of CoQ(10) in a cell having endogenous DDB  
CC activity. This sequence represents a protein relating to the isoprenoid  
CC production of the invention.

XX Sequence 388 AA;

Query Match 50.4%; Score 999.5; DB 23; Length 388;  
Best Local Similarity 50.9%; Pred. No. 5,6e-89;  
Matches 199; Conservative 74; Mismatches 109; Indels 9; Gaps 3;

QY 4 ICILGATGSIQVSTLDVVARHPPDKYQVVALTANGNIDALYEQCLAHPREYAVVMSKVA 63  
DB 7 VVVLGATGSIQVSTLDVVARHPPDKYQVVALTANGNIDALYEQCLAHPREYAVVMSKVA 66  
QY 64 EFKQRIASPVADIKVLGSEALQOVATLENDVTMAAIVGAAGLPTLAAKAGTVLL 123  
DB 67 DIKEALAGS---SVEAAGADALVEAAMM-GADMTMAALIGCAGLKATLAIRKGTVAL 122  
QY 124 ANKEALVMSGOIFMOAVSDSGAVLLPIDSEHNAIFQCKPAGYTPGHTAKQARRILLTASG 183  
DB 123 ANKESLVSKGLMIDVAREHGTLLPVDEHNAIFQCF----PHARDYVRRIITTSAG 177  
QY 184 GPFRTPIETLSSVTPDQAAVAPKQDMGKIKISVDSATWNNKGLLEIACCLFNMEDQIE 243  
DB 178 GPFRTTSLAEMATVPERAVOHPNMGAKISIDSATWNNKGLLEIACCLFNMEDQIE 237  
QY 244 VVIHQSIHSHWVDVYDGSVLQMGNDPDRTPIAHAMAFERPDGVAFLDIEFVGHMDF 303  
DB 238 IIVHPOSVTHSWVEYDGSILQIGSPDRTPIGHILAMPKMETPAESLDFTKLRQMDP 297  
QY 304 EKPDLRPFCLRIAYEAIKSGGIMPTVLNANEIAVEAFINEEVKFTDIAVILERSMAQF 363  
DB 298 BAPDYERFPALTLAMESIKSGGARPAVMAANEIAVAAPFLDKKIGFLDIKIVEKTLDRY 357  
QY 364 KPDDAGSLLEYLQADQDAREVARDIKTLVA 394  
DB 358 TPATPSLIEDVAFIDNEARIQAALMESLPA 388

Search completed: January 29, 2004, 15:49:39  
Job time : 37.0149 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 29, 2004, 15:45:34 ; Search time 12.3656 Seconds  
(without alignments)  
1348.130 Million cell updates/sec

Title: US-09-941-947a-8

Perfect score: 1985  
Sequence: 1 MKGICILGATSGVSTLDV.....IQADQDAEVARDDIKTLVA 394

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/6D.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description          |
|------------|--------|-------------|--------|-------|----------------------|
| 1          | 1215.5 | 61.2        | 617    | 4     | US-09-252-991A-21113 |
| 2          | 1114.5 | 56.1        | 398    | 4     | US-09-328-352-7199   |
| 3          | 818.5  | 41.2        | 399    | 4     | US-09-491-362-7      |
| 4          | 818.5  | 41.2        | 399    | 4     | US-09-874-562-7      |
| 5          | 818.5  | 41.2        | 477    | 4     | US-09-449-335-6      |
| 6          | 817.5  | 41.2        | 477    | 4     | US-09-480-921B-10    |
| 7          | 814.5  | 41.0        | 477    | 4     | US-09-449-335-2      |
| 8          | 813.5  | 41.0        | 477    | 4     | US-09-480-921B-29    |
| 9          | 741    | 37.3        | 475    | 3     | US-09-874-562-2      |
| 10         | 741    | 37.3        | 475    | 3     | US-09-874-562-2      |
| 11         | 620    | 31.2        | 340    | 4     | US-09-198-452A-362   |
| 12         | 111    | 5.6         | 671    | 4     | US-09-252-991A-1862  |
| 13         | 107    | 5.4         | 2285   | 4     | US-09-252-991A-17790 |
| 14         | 105.5  | 5.3         | 3816   | 4     | US-09-428-517-3      |
| 15         | 99     | 5.0         | 348    | 4     | US-09-252-991A-20932 |
| 16         | 99     | 5.0         | 348    | 4     | US-09-252-991A-17607 |
| 17         | 98.5   | 5.0         | 390    | 4     | US-09-328-352-4916   |
| 18         | 97     | 4.9         | 292    | 4     | US-09-501-115-28     |
| 19         | 96.5   | 4.9         | 805    | 4     | US-09-252-991A-32752 |
| 20         | 96     | 4.8         | 423    | 4     | US-09-328-352-5657   |
| 21         | 95     | 4.8         | 1049   | 4     | US-09-252-991A-25336 |
| 22         | 94.5   | 4.8         | 319    | 4     | US-09-252-991A-20187 |
| 23         | 94.5   | 4.8         | 425    | 4     | US-09-134-001C-4433  |
| 24         | 94     | 4.7         | 461    | 3     | US-09-355-115-7      |
| 25         | 94     | 4.7         | 568    | 4     | US-09-252-991A-22727 |
| 26         | 93.5   | 4.7         | 581    | 4     | US-09-252-991A-25556 |
| 27         | 93.5   | 4.7         | 628    | 3     | US-08-776-271-2      |

|    |      |     |       |   |                      |                    |
|----|------|-----|-------|---|----------------------|--------------------|
| 28 | 93.5 | 4.7 | 628   | 3 | US-09-215-035-2      | Sequence 2, Appli  |
| 29 | 93   | 4.7 | 1346  | 3 | US-09-105-537-37     | Sequence 37, Appl  |
| 30 | 93   | 4.7 | 11877 | 3 | US-09-105-537-6      | Sequence 6, Appli  |
| 31 | 92.5 | 4.7 | 1562  | 3 | US-09-320-878-3      | Sequence 3, Appli  |
| 32 | 92.5 | 4.7 | 1562  | 3 | US-09-105-537-35     | Sequence 35, Appl  |
| 33 | 92.5 | 4.7 | 1562  | 4 | US-09-141-908-4      | Sequence 4, Appli  |
| 34 | 92.5 | 4.7 | 1562  | 4 | US-09-657-440-3      | Sequence 3, Appli  |
| 35 | 91.5 | 4.6 | 584   | 1 | US-08-426-819A-36    | Sequence 36, Appl  |
| 36 | 91.5 | 4.6 | 622   | 1 | US-08-426-819A-35    | Sequence 35, Appl  |
| 37 | 91.5 | 4.6 | 673   | 4 | US-09-328-352-5834   | Sequence 5834, Ap  |
| 38 | 91.5 | 4.6 | 710   | 4 | US-09-252-991A-32789 | Sequence 32789, A  |
| 39 | 91   | 4.6 | 1346  | 3 | US-09-320-878-4      | Sequence 4, Appli  |
| 40 | 91   | 4.6 | 1346  | 4 | US-09-141-908-5      | Sequence 5, Appli  |
| 41 | 91   | 4.6 | 1346  | 4 | US-09-657-440-4      | Sequence 4, Appli  |
| 42 | 90.5 | 4.6 | 752   | 4 | US-09-252-991A-29967 | Sequence 29967, A  |
| 43 | 90   | 4.5 | 695   | 6 | 5460961-5            | Patent No. 5460961 |
| 44 | 89.5 | 4.5 | 706   | 4 | US-09-252-991A-25730 | Sequence 25730, A  |
| 45 | 89.5 | 4.5 | 6396  | 4 | US-09-410-551B-72    | Sequence 72, Appl  |

## ALIGNMENTS

RESULT 1  
US-09-252-991A-21113  
Sequence 21113, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 21113  
LENGTH: 617  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21113

| Query Match | Best Local Similarity  | Score         | DB 4     | Length | 617 |
|-------------|--|---------------|----------|--------|-----|
| Matches 246 | Conservative 47  | Mismatches 88 | Indels 1 | Gaps 1 |     |
| 4           | ICIGATSGISGVSTLDVVAHHPDKYQVVALTANGNIDALVEQCLAHHPYAVVMMESKVA  | 63            |          |        |     |
| 228         | ISVIGATSGISGVSTLDVVAHHPDKYQVVALTANGNIDALVEQCLAHHPYAVVMMESKVA | 287           |          |        |     |
| 64          | EFKQRIASPVADIKVTSSEBALQVATLENDVTMAIYGAAGLLPTLAAKRGKTYLL      | 123           |          |        |     |
| 288         | ALQGLIAAGR-TRVIFGEOALCEVASAEVDMVAIYGAAGLLPTLAAKRGKTYLL       | 346           |          |        |     |
| 124         | ANKALVMSQIFPQVAVSDGAVLLPDSSENAIFQCPAGYTGHTAKQARRLLTASG       | 183           |          |        |     |
| 347         | ANKALVMSQIFPQVAVSDGAVLLPDSSENAIFQCPAGYTGHTAKQARRLLTASG       | 406           |          |        |     |
| 184         | GPFRPTIFLTSSVTPDAVAHPKMGKRI SVDSATVANKGLEILACLIENKPDQIE      | 243           |          |        |     |
| 407         | GPFRPTIFLTSSVTPDAVAHPKMGKRI SVDSATVANKGLEILACLIENKPDQIE      | 466           |          |        |     |
| 244         | VVHPOSHISMVDYVGVSVLAQWGNPDMPPIAHAAWPERPDGVAPLDIFVGHDF        | 303           |          |        |     |
| 467         | VVHPOSHISMVDYVGVSVLAQWGNPDMPPIAHAAWPERPDGVAPLDIFVGHDF        | 526           |          |        |     |
| 304         | EXPDLPKFPCLRLAYEAIKSGIMPTVLNANELAYEAFINEYKPTDIAVILERSNAQF    | 363           |          |        |     |
| 527         | ORPDEQFPCLRLAYEAIKSGIMPTVLNANELAYEAFINEYKPTDIAVILERSNAQF     | 586           |          |        |     |

QY 364 KPDDAGSLVLTQADQAREVA 385  
 DB 587 AVTAVESLDQVLAADRARSVA 608

## RESULT 2

US-09-328-352-7199  
 ; Sequence 7199, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 7199  
 ; LENGTH: 398  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 ; US-09-328-352-7199

Query Match 56.1%; Score 1114.5; DB 4; Length 398;  
 Best Local Similarity 56.7%; Pred. No. 3.4e-114;  
 Matches 224; Conservative 58; Mismatches 108; Indels 5; Gaps 2;

QY 2 KGICILGATGSGVTLDDVVAHHPDKYQVVALTANGNIDALTYEQLAHPEYAVVNESK 61  
 DB 3 QSVCLIGVTVGSGRSTLTKLIGHPKYSVFVSAHSRISSEVEICKOPRPK-VVVVPEOK 61  
 QY 62 VAEFKRIASPVADIKVLSGSEALQOVATLENVDVTMAIYGAAGLPTLAAAKGKTIV 121  
 DB 62 IAEIKTLPAQONISIDIVLAGBGLVDIASHTDVIWMAIVGAAGLPTLAAVAGKRV 121  
 QY 122 LIANKEALVMSGOIFMQAVSDSGAVLPLIDSEHNAIPQCPAPYTPGHTAKQAR---RI 177  
 DB 122 LIANKEALVMSGEIMVQARBEHQALLPVDSEHNAIPQSLPANYLQADRTGQPOLGVSKI 181  
 QY 178 LIATAGGPRRTPIETLSVTPDOAVAHKMDGKISVDSATMNNKGLLEIECLTFNM 237  
 DB 182 LIATAGGPRPLNLSHQLTHVTYQACKHNMWNGQIISVDSATLNNKGLLEIECHLPSI 241  
 QY 238 BPDQEVVTHPQSIHSMVYVDSVLAQMGNDPMTPIAHAAWPERFDSGVAPLDFE 297  
 DB 242 SEHFTVAVHPQSIHSMVYVDSVLAQMGNDPMTPIAHAAWPERLQTVVPALDFE 301  
 QY 298 VEHDPFKPDLKRPCLALAYEALIKSGGIMPTVLNANEIYEAFLNEAVKFTDIAYIE 357  
 DB 302 YSQLFQADDTQKFPALNINARQMRAGGLAPILNANEIYEAFLMERIGFTSIPYVE 361  
 QY 358 RMAQFKPDDAGSLVLTQADQAREVARDIKTL 392  
 DB 362 HTLEKLENNAAASIEICILDKQVAAVSVAQYIISI 396

RESULT 3  
 US-09-491-362-7  
 ; Sequence 7, Application US/09491362  
 ; Patent No. 6281017  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Croteau, Rodney B  
 ; APPLICANT: Lange, Bernd M  
 ; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND  
 ; FILE REFERENCE: WSIU14977  
 ; CURRENT APPLICATION NUMBER: US/09/491,362  
 ; CURRENT FILING DATE: 2000-01-26  
 ; EARLIER APPLICATION NUMBER: 60/118,349  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 7

QY 364 KPDDAGSLVLTQADQAREVA 385  
 DB 587 AVTAVESLDQVLAADRARSVA 608

LENGTH: 399  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-491-362-7

Query Match 41.2%; Score 818.5; DB 3; Length 399;  
 Best Local Similarity 44.6%; Pred. No. 1.6e-81;  
 Matches 176; Conservative 71; Mismatches 131; Indels 17; Gaps 5;

QY 2 KGICILGATGSGVTLDDVVAHHPDKYQVVALTANGNIDALTYEQLAHPEYAVVNESK 61  
 DB 3 KPISIVSTGSIQHTIDIVAEKDPKRVVLAAGSVTLTADQVRRFKALVAVERESL 62  
 QY 62 VAEFKRIASPVADIKVLSGSEALQOVATLENVDVTMAIYGAAGLPTLAAAKGKTIV 121  
 DB 63 INELKEALADLDY-KLIIIPGEGVIEVAHHPAIVTVTIVGAGLKPVAALIEAGKDI 121  
 QY 122 LIANKEALVMSGOIFMQAVSDSGAVLPLIDSEHNAIPQCPAPYTPGHTAKQARIL 178  
 DB 122 ALANKETLIAGGPRVPLFANKHNVKILPADSEHSAIPQCIQGPBG-----ALRKII 173  
 QY 179 LTAGGPRRTPIETLSVTPDOAVAHKMDGKISVDSATMNNKGLLEIECLTFNME 238  
 DB 174 LTAGGAFRPMVPEKLEKVVADALKHPNMWNGKLTVDATLFNKLEIYEAHYLGAE 233  
 QY 239 PDIETVTHPQSIHSMVYVDSVLAQMGNDPMTPIAHAAWPERFDSGVAP---LDI 295  
 DB 234 YDIEIYTHPQSIHSMVYVDSVLAQMGNDPMTPIAHAAWPERFDSGVAP---LDI 293  
 QY 296 FEVHDPFKPDLKRPCLALAYEALIKSGGIMPTVLNANEIYEAFLNEAVKFTDIAYI 355  
 DB 294 CKLGSLFFKPKPDNKKYPSMDLAVYAGAGGTMTGVSAANEKAVEMFIDEXISYLDIFKV 353  
 QY 356 IERSMAQFKPD--DAGSLVLTQADQAREVARDI 388  
 DB 354 VELTCDHRNELVTSPSLBIYHYDLMARBYAANV 388

RESULT 4  
 US-09-874-562-7  
 ; Sequence 7, Application US/09874562  
 ; Patent No. 6420159  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Croteau, Rodney B  
 ; APPLICANT: Lange, Bernd M  
 ; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND  
 ; FILE REFERENCE: WSIU17549  
 ; CURRENT APPLICATION NUMBER: US/09/874,562  
 ; CURRENT FILING DATE: 2001-06-04  
 ; PRIOR APPLICATION NUMBER: 09/491,362  
 ; PRIOR FILING DATE: 2000-01-26  
 ; PRIOR APPLICATION NUMBER: 60/118,349  
 ; PRIOR FILING DATE: 1999-02-03  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 7  
 ; LENGTH: 399  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-874-562-7

QY 364 KPDDAGSLVLTQADQAREVA 385  
 DB 587 AVTAVESLDQVLAADRARSVA 608

QY 122 LLANKEALVMSGQIFMOWSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARRIL 178  
 DB 122 ALANKETLJAGSPFPLPLANKHNVKILPADSEHSAIFQCIQGLPBG-----ALRKII 173  
 QY 179 LTAAGSPRRRTIEFLSSVTPDOAVAHPRKMDKRIKISVDSATMNGKIELLEIACILFPMWE 238  
 DB 174 LTAAGSARFDMWEVETLKEKVKADALKHNNMNGKILTVDSATLFRKGLIEVIEAHYLFQAE 233  
 QY 239 PDOIIEVVIHPQSIHSMVDYVDGSVLAQMGNDPMTPIAHAMAPERPDSGVAP---LDI 295  
 DB 234 YDDIIEVVIHPQSIHSMIFQOSSVLAQMGNDPMTPIAHAMAPERPDSGVAP---LDI 293  
 QY 236 FEVGMDEKPKDPKRPCLRLAYEAKISGIMPTVLNANEIYAEAFINEEVKFTDIAVI 355  
 DB 294 CKLGSLEFKKPDNVKYPKMDLAVYAGAGGTMTGVLISANEKAVEMFIDEKISYDIPKV 353  
 QY 356 IERSMAQFKPD--DAGSLELVLAQDQDAREVARDI 388  
 DB 354 VELTCDKHNEELVTSPLSEIIVHYDLMAREYANV 388

RESULT 5  
 US-09-449-335-6  
 Sequence 6, Application US/09449335

GENERAL INFORMATION:  
 APPLICANT: Busch, Marco  
 APPLICANT: Hain, Rüdiger  
 APPLICANT: Martin, William  
 APPLICANT: Tietjen, Klaus  
 APPLICANT: Klotz, Andreas  
 TITLE OF INVENTION: Method of determining the activity of  
 TITLE OF INVENTION: 1-deoxy-D-xylulose-5-phosphate reductoisomerase and  
 FILE REFERENCE: 2020US  
 CURRENT APPLICATION NUMBER: US/09/449,335  
 EARLIER FILING DATE: 1999-11-24  
 EARLIER APPLICATION NUMBER: DE 199 35 967.9  
 EARLIER FILING DATE: 1999-07-30  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: Patent In Ver. 2.1  
 SEQ ID NO 6  
 LENGTH: 477  
 TYPE: PRF  
 ORGANISM: Arabidopsis thaliana  
 US-09-449-335-6

Query Match 41.2%; Score 818.5; DB 4; Length 477;

Best Local Similarity 44.6%; Pred. No. 2,2e-81; Matches 176; Conservative 71; Mismatches 131; Indels 17; Gaps 5;

QY 2 KICILGATGSGVSTLDVVAHPRKQVVALTANGINDALYEOCLAHPRVAVVWESK 61  
 DB 81 KRISIVSGTSGIGTLDIIVAEHPKFRVVALAAGSVNTLLADQVRFRKPAIVAVRNSL 140  
 QY 62 VAEFKRIASPVADI---KVLGSEALQOVATLENVPMMAIYGAAGLPTIAAARAKTV 121  
 DB 141 INELKEALADLVY-KLEIIPGEGVTEVARHPEAVTVTGIVGAGLKPVTAAIEAGDI 199  
 QY 122 LLANKEALVMSGQIFMOWSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARRIL 178  
 DB 200 ALANKETLJAGSPFPLPLANKHNVKILPADSEHSAIFQCIQGLPBG-----ALRKII 251  
 QY 179 LTAAGSPRRRTIEFLSSVTPDOAVAHPRKMDKRIKISVDSATMNGKIELLEIACILFPMWE 238  
 DB 252 LTAAGSARFDMWEVETLKEKVKADALKHNNMNGKILTVDSATLFRKGLIEVIEAHYLFQAE 311  
 QY 239 PDOIIEVVIHPQSIHSMVDYVDGSVLAQMGNDPMTPIAHAMAPERPDSGVAP---LDI 295  
 DB 312 YDDIIEVVIHPQSIHSMIFQOSSVLAQMGNDPMTPIAHAMAPERPDSGVAP---LDI 371  
 QY 356 FEVGMDEKPKDPKRPCLRLAYEAKISGIMPTVLNANEIYAEAFINEEVKFTDIAVI 355

DB 372 CKLGSLEFKKPDNVKYPKMDLAVYAGAGGTMTGVLISANEKAVEMFIDEKISYDIPKV 431  
 QY 356 IERSMAQFKPD--DAGSLELVLAQDQDAREVARDI 388  
 DB 432 VELTCDKHNEELVTSPLSEIIVHYDLMAREYANV 466

RESULT 6  
 US-09-480-921B-10  
 Sequence 10, Application US/09480921B

GENERAL INFORMATION:  
 APPLICANT: Levin, Joshua Z.  
 APPLICANT: Budizewski, Gregory J.  
 APPLICANT: Potter, Sharon L.  
 APPLICANT: Megrich, Lynette M.  
 TITLE OF INVENTION: Herbicide Target Genes and Methods  
 FILE REFERENCE: PB/5-30780A  
 CURRENT APPLICATION NUMBER: US/09/480,921B  
 EARLIER FILING DATE: 2000-01-11  
 NUMBER OF SEQ ID NOS: 29  
 SOFTWARE: Patent In Ver. 2.1  
 SEQ ID NO 10  
 LENGTH: 477  
 TYPE: PRF  
 ORGANISM: Arabidopsis thaliana  
 US-09-480-921B-10

Query Match 41.2%; Score 817.5; DB 4; Length 477;  
 Best Local Similarity 44.6%; Pred. No. 2,2e-81; Matches 178; Conservative 69; Mismatches 127; Indels 25; Gaps 6;

QY 2 KICILGATGSGVSTLDVVAHPRKQVVALTANGINDALYEOCLAHPRVAVVWESK 61  
 DB 81 KRISIVSGTSGIGTLDIIVAEHPKFRVVALAAGSVNTLLADQVRFRKPAIVAVRNSL 140  
 QY 62 VAEFKRIASPVADI---KVLGSEALQOVATLENVPMMAIYGAAGLPTIAAARAKTV 117  
 DB 141 INELKEALADLVY-KLEIIPGEGVTEVARHPEAVTVTGIVGAGLKPVTAAIEAGDI 195  
 QY 118 KTVLANKREALVMSGQIFMOWSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARRIL 174  
 DB 196 KTVLANKREALVMSGQIFMOWSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARRIL 247  
 QY 175 KRILITASGSPRRRTIEFLSSVTPDOAVAHPRKMDKRIKISVDSATMNGKIELLEIACIL 234  
 DB 248 KRILITASGSPRRRTIEFLSSVTPDOAVAHPRKMDKRIKISVDSATMNGKIELLEIACIL 307  
 QY 235 FMEPPDOIIEVVIHPQSIHSMVDYVDGSVLAQMGNDPMTPIAHAMAPERPDSGVAP---LDI 292  
 DB 308 FMEPPDOIIEVVIHPQSIHSMIFQOSSVLAQMGNDPMTPIAHAMAPERPDSGVAP---LDI 367  
 QY 293 LDIIEVGMDEKPKDPKRPCLRLAYEAKISGIMPTVLNANEIYAEAFINEEVKFTDIAVI 351  
 DB 368 LDIIEVGMDEKPKDPKRPCLRLAYEAKISGIMPTVLNANEIYAEAFINEEVKFTDIAVI 427  
 QY 352 IAVIERSMAQFKPD--DAGSLELVLAQDQDAREVARDI 388  
 DB 428 IAVIERSMAQFKPD--DAGSLELVLAQDQDAREVARDI 466

RESULT 7  
 US-09-449-335-2

Sequence 2, Application US/09449335

GENERAL INFORMATION:  
 APPLICANT: Busch, Marco  
 APPLICANT: Hain, Rüdiger  
 APPLICANT: Martin, William  
 APPLICANT: Tietjen, Klaus  
 APPLICANT: Klotz, Andreas  
 TITLE OF INVENTION: Method of determining the activity of



TITLE OF INVENTION: 1-deoxy-D-xyloside-5-phosphate reductoisomerase and  
 TITLE OF INVENTION: 1-deoxy-D-xyloside-5-phosphate synthase  
 FILE REFERENCE: 2020US  
 CURRENT APPLICATION NUMBER: US/09/449,335  
 CURRENT FILING DATE: 1999-11-24  
 EARLIER APPLICATION NUMBER: DB 199 35 967.9  
 EARLIER FILING DATE: 1999-07-30  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 2  
 LENGTH: 477  
 TYPE: PRT  
 ORGANISM: Arabidopsis thaliana  
 US-09-449-335-2

Query Match 41.0%; Score 814.5; DB 4; Length 477;  
 Best Local Similarity 44.3%; Pred. No. 6,1e-81;  
 Matches 175; Conservative 71; Mismatches 132; Indels 17; Gaps 5;

QY 2 KGCILGATGSGVSTLDVVAHHPKQVVALTANGIDALYEQCLAHHEPVAVVMESK 61  
 DB 81 KPISIVSGSIGTGTLDIVAENPDKFRVALAAGSNVTLADQVRRFRKPAVAVRNSL 140  
 QY 62 VAEFKRIASPVADIKVLSSGEALQOVATLENVDTMAAIVGAGLPTLAAAGKTV 121  
 DB 141 INELKRALADLDY-KLEIIRGQGVIEVARHPEAVTVTGIVGAGLPTVAALIEAGKDI 199  
 QY 122 LIANTBALVMSQIIMQAVSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQARIL 178  
 DB 200 ALANKETILAGPFVPLPANKNNVXILLPADSEHSAIFOCIQGLPEG-----ALRKII 251  
 QY 179 LTAAGGPPRRPIETLSSVTPDQAVAHFPKDGKISVDSATMNMKGLELEACLLFNM 238  
 DB 252 LTAAGGAFRDMVEKLEKVKADALKHPNMNMKGKITVDSCTLPMKGLVIEHAHLFPAE 311  
 QY 239 PDQIEVVIHPQSIHSMVDYVDSVLAQMGNDMPKTPIAHAAWBERFDSGVAP--LDI 295  
 DB 312 YDIEIVIHPOSIHSMITQDSVLAQMGWPMKPLITYTMSWDVRVPCSEVTWPRDL 371  
 QY 296 FEVGMDFEKPDKFPCCLRLAYEAIKSGIMPTVLAANEIAVAEFLNEVKFTDIAVI 355  
 DB 372 CKLGSLTFKKPDVVKYPSMDLAVAAAGAGTGTGVAANERKAVEMFIDEKISYLDIKV 431  
 QY 356 IERSMAQFKPD--DAGSLEVLQADQDAFEVARDI 388  
 DB 433 VZLTCDKHRELVTSPSLEIIVHYDLMAEVAAYV 466

RESULT 8  
 US-09-480-921B-29  
 Sequence 29, Application US/09480921B  
 Patent No. 6387637  
 GENERAL INFORMATION:  
 APPLICANT: Levitt, Joshua Z.  
 APPLICANT: Budziszewski, Gregory J.  
 APPLICANT: Potter, Sharon L.  
 APPLICANT: Weglich, Lynette M.  
 TITLE OF INVENTION: Herbicide Target Genes and Methods  
 FILE REFERENCE: PB/5-30780A  
 CURRENT APPLICATION NUMBER: US/09/480,921B  
 NUMBER OF SEQ ID NOS: 29  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 29  
 LENGTH: 477  
 TYPE: PRT  
 ORGANISM: Arabidopsis thaliana  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (39)  
 OTHER INFORMATION: Xaa = Asp or Asn  
 NAME/KEY: SITE  
 LOCATION: (155)

OTHER INFORMATION: Xaa = Pro or Leu  
 NAME/KEY: SITE  
 LOCATION: (162)  
 OTHER INFORMATION: Xaa = Leu or Gln  
 NAME/KEY: SITE  
 LOCATION: (187)  
 OTHER INFORMATION: Xaa = Lys or Gln  
 NAME/KEY: SITE  
 LOCATION: (465)  
 OTHER INFORMATION: Xaa = Asp or Asn  
 US-09-480-921B-29

Query Match 41.0%; Score 813.5; DB 4; Length 477;  
 Best Local Similarity 44.4%; Pred. No. 7.8e-81;  
 Matches 177; Conservative 69; Mismatches 128; Indels 25; Gaps 6;

QY 2 KGCILGATGSGVSTLDVVAHHPKQVVALTANGIDALYEQCLAHHEPVAVVMESK 61  
 DB 81 KPISIVSGSIGTGTLDIVAENPDKFRVALAAGSNVTLADQVRRFRKPAVAVRNSL 140  
 QY 62 VAEFKRIASPVADI---KVLSSGEALQOVATLENVDTMAAIVGAGLPTLAAAKA 117  
 DB 141 INELKEAL-----ADLYKXELIPGEXGVIEVARHPEAVTVTGIVGAGLPTVAALIEA 195  
 QY 118 GKVTLANKALVMSQIIMQAVSDGAVLLPIDSEHNAIFOC---MPAGYTPGHTAKQA 174  
 DB 196 GKDIALANKETILAGPFVPLPANKNNVXILLPADSEHSAIFOCIQGLPEG-----AL 247  
 QY 175 RLILTAAGGPPRRPIETLSSVTPDQAVAHFPKDGKISVDSATMNMKGLELEACLL 234  
 DB 248 RKILITASGAFRDMPEKLEKVKADALKHPNMNMKGKITVDSATLFNGLVIEHAHL 307  
 QY 235 FNMPPDQIEVVIHPQSIHSMVDYVDSVLAQMGNDMPKTPIAHAAWBERFDSGVAP-- 292  
 DB 308 PDAEYDIEIVIHPOSIHSMITQDSVLAQMGWPMKPLITYTMSWDVRVPCSEVTW 367  
 QY 293 -LDIFEVGMDFEKPDKFPCCLRLAYEAIKSGIMPTVLAANEIAVAEFLNEVKFTD 351  
 DB 368 RLDCIKLSLTFKKPDVVKYPSMDLAVAAAGAGTGTGVAANERKAVEMFIDEKISYLD 427  
 QY 352 IAVIERSMAQFKPD--DAGSLEVLQADQDAFEVARDI 388  
 DB 428 IFKVBELTCDKHRELVTSPSLEIIVHYDLMAEVAAYV 466

RESULT 9  
 US-09-491-362-2  
 Sequence 2, Application US/09491362  
 Patent No. 6281017  
 GENERAL INFORMATION:  
 APPLICANT: Croteau, Rodney B  
 APPLICANT: Lange, Bernd M  
 TITLE OF INVENTION: 1-DEOXY-D-XYLOSIDE-5-PHOSPHATE REDUCTOISOMERASE, AND  
 TITLE OF INVENTION: METHODS OF USE  
 FILE REFERENCE: NSUR14972  
 CURRENT APPLICATION NUMBER: US/09/491,362  
 CURRENT FILING DATE: 2000-01-26  
 EARLIER APPLICATION NUMBER: 60/118,349  
 EARLIER FILING DATE: 1999-02-03  
 NUMBER OF SEQ ID NOS: 13  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 2  
 LENGTH: 475  
 TYPE: PRT  
 ORGANISM: Mentha piperita  
 US-09-491-362-2

Query Match 37.3%; Score 741; DB 3; Length 475;  
 Best Local Similarity 43.0%; Pred. No. 7.8e-73;  
 Matches 172; Conservative 73; Mismatches 133; Indels 22; Gaps 10;  
 QY 2 KGCILGATGSGVSTLDVVAHHPKQVVALTANGIDALYEQCLAHHEPVAVVMESK 61  
 DB 81 KPISIVSGSIGTGTLDIVAENPDKFRVALAAGSNVTLADQVRRFRKPAVAVRNSL 140



PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 31862  
 LENGTH: 671  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-31862

Query Match 5.6%; Score 111; DB 4; Length 671;  
 Best Local Similarity 21.1%; Pred. No. 0.005;  
 Matches 85; Conservative 58; Mismatches 139; Indels 120; Gaps 20;

QY 9 ATGSGVSTLDVVAHPDKYQVVALTANGNIDALYEQ---CLAHPEYAVVMES--K 61  
 DB 302 AAGLVVAATLDLAAH-----AAGDHAIIIEBQPVGQRAAEVIGLNAAYE 351  
 QY 62 VAEFKRIASPVADIKVLSGSEALQO-----VATLENDVTMAAIVGAAG 107  
 DB 352 AAEVHQRNRERGLVEVARLHPGEGIDHAQPFVQAVGAERPVVAERAGVEVAQAADLGIG 411  
 QY 108 -LLPTLAAKAKTLLANKKALVMSGOIFMOAVDSG-----AVLLPDSHNAIF 158  
 DB 412 AAVAAACQAVDDIALVADEDAVGDFVVALAADAAGERRQVGEVLAAND--HGVL 469  
 QY 159 -----QCM-----PAGYTPGHTAQARRIILITASGGERPRTPIFTLSV 197  
 DB 470 LVQVGIQAQEAACLRKEMGVGSGRAGQORAHHAQAFLIVQVAG-----QAQLA 521  
 QY 198 TPDOAVA-HPKV-----DMGRKISVDSATMMKGLIELICLLFMEDQIEV-----VI 246  
 DB 522 TEVQAGHSGKLAALMTDGGIIVADVAVAAHAGVLAER-----PADVEVSLELAAL 574  
 QY 247 HP-----QSIHNSVVDYDSV-LAQMNDPKMETPIAHAMAMBERPDSGVAPLDIEF- 297  
 DB 575 GPIDADAGQLVYGAAGHVQDGAADAAAGNP-----AEQVWMSLEHFA 619  
 QY 298 VEHMDEKEDLKRFPCLRLAYEAKISGGIMPTVLAANETIAV 339  
 DB 620 LDEFPDPPERQO-----AISTVE-GDVGVLAHBAANVAV 654

RESULT 13  
 US-09-252-991A-17790  
 Sequence 17790, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 17790  
 LENGTH: 2285  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-17790

Query Match 5.4%; Score 107; DB 4; Length 2285;  
 Best Local Similarity 22.8%; Pred. No. 0.11;  
 Matches 87; Conservative 51; Mismatches 137; Indels 106; Gaps 19;

QY 6 ILGATGSGVSTLDVVAHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMESKVAEF 65  
 DB 1505 LIGAT--FVELLMRLAEHPDAADFAFLAPRR-DA-----ASQPEPLVDV---VALF 1550  
 QY 66 KORIASPVADIKVLSGSEAL--QQVATLENDVTMAAIVGAAGLLPTLAAAKAKTVL 122

DB 1551 ERQVEALP-----GSAALAFEBQKWTYRDLDHVARCV--ATRLV--RAGARRGDAIG 1598  
 QY 123 LANKKALVMSGOIFMOAVSDSGAVLLPDSHNAIFQCMFAGTTPGHTAQARRIILITAS 182  
 DB 1599 VALNRSPEMATATM--GILRAGIVCVPLDVSYP-----QRTALLLET 1639  
 QY 183 GGPFRRT-----PIETLSVTPDOAVAHPKMD-MGRKISVDSATMMKGL 226  
 DB 1640 AQPFRVVAHPERHVAALAEKVLPEVEHVAIDIKETFAAPQDLDELAMLLFTSGSGRRKGV 1699  
 QY 227 ELIEACLLFNNEDQIEVVIHPQSIHSM--VDYVDSVLAQMGNPDRT---PIAHAMA 281  
 DB 1700 EL-----SHRMANYTQMQLRVAASGVPGGLRTLPAPLSPDMA 1736  
 QY 282 WPERFDS--GVAPLD--IFEVGHMD-----FEKPDLR---FPCLRLAYEAKISGGIM 327  
 DB 1737 FOELFTLCGGEGEQLQILSNRERDPSALLHYLERQVRLPVALQRLAENSALGVR 1796  
 QY 328 PTVLAANEIAVNAEAFINEEYK 348  
 DB 1797 FGALRVVVSSEQLRTIEDVR 1817

RESULT 14  
 US-09-428-517-3  
 Sequence 3, Application US/09428517  
 Patent No. 6251636  
 GENERAL INFORMATION:  
 APPLICANT: Betlach, Mary C.  
 APPLICANT: Shah, Sanjay Krishnakant  
 APPLICANT: McDaniel, Robert  
 TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
 FILE REFERENCE: 30062-20029.00  
 CURRENT APPLICATION NUMBER: US/09/428,517  
 CURRENT FILING DATE: 1999-10-28  
 EARLIER APPLICATION NUMBER: 60/120,254  
 EARLIER FILING DATE: 1999-02-16  
 EARLIER APPLICATION NUMBER: 60/106,100  
 EARLIER FILING DATE: 1998-10-29  
 NUMBER OF SEQ ID NOS: 12  
 SOFTWARE: Patent In Ver. 2.1  
 SEQ ID NO 3  
 LENGTH: 3816  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: Recombinant  
 OTHER INFORMATION: Oleandolide PKS  
 US-09-428-517-3

Query Match 5.3%; Score 105.5; DB 3; Length 3816;  
 Best Local Similarity 17.8%; Pred. No. 0.41;  
 Matches 98; Conservative 76; Mismatches 151; Indels 227; Gaps 25;

QY 2 KGICHLATGSGVSTLDVVAHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMESK 61  
 DB 3163 ESTVYHSAAGVGMAAVQ-LAKHMDA-EVFGTAKSGKMDVLAAGLDBER-----ISSR 3215  
 QY 62 VAEFKRIASPVADIKVLSGSEALQQVATLENDVTMAA-----101  
 DB 3216 TTEFEQCFRAT-----SGSGGIDIVATNALSGDFDASARLREBGRFVEMKTDIRT 3267  
 QY 102 ---TVGAAGLLPTLA-----AAKAGKTVLLANKKALVMSGOIF--MOAVSDSGAVLL-- 148  
 DB 3268 DLGVVAGDG--VPDIRYAFDLAENGAERI-----QQMLDEIALLTDAGVLRPL 3316  
 QY 149 ---PIDSEHNAIFQCMFAGTTPGHTAQARR-----117ASG-----183  
 DB 3317 RAMPVRBAHAL-----RFVQAHRHVGVVLTVPALDABGTVALITGAGTICAL 3365  
 QY 184 -----GPRFRPTITLSVTPDOAVAHPKMDMKKISVDSATMMN-KGLB-----227

Db 3366 VARRVTHDVTRRLVSRSGVAPD-LAELGALCAEYTAACDVANRKAALKALBDIP 3423  
 QY 228 -----LIEACLTFNMEPDQIEVYTHPQ-----SIHSMVDYVD-----G 261  
 Db 3424 PEHVPTGIHVHAGVLDGVDGVSGLTFPERVDYVTKKPVDAALTLIESVIGSLDDPALFVIFS 3483  
 QY 262 SYLAQMGNDP-----MRTPIAH-----AMW-----PER 285  
 Db 3484 SAASKVLGGGGGSAANQFIDTLARHARRGRTSVSLGCMGIMHESGLTGGLEADIDRDR 3543  
 QY 286 FD-SGVAPLDIFEVGHM-----DFEKPDLKFPFCRLAYEAKSGGIMPTVYN----- 332  
 Db 3544 MSBAGIAMPIDBAHLFDRALELDGDPVLLPRLNEALBDBAAGCTPPLISGLVVRH 3603  
 QY 333 -----AANETAEAPLNEEYKFTDIAVIERSMAGFPDDAGSLLEVLAQDQDAR 382  
 Db 3604 RPSAPAGTATAPATGPFAPFARE-----LAAAPDPR 3634  
 QY 363 EVARDIITLVVA 394  
 Db 3635 RALRDLVGRHVA 3646

RESULT 15  
 US-09-252-991A-20932  
 ; Sequence 20932, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 20932  
 ; LENGTH: 482  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-20932

Query Match 5.0%; Score 99; DB 4; Length 482;  
 Best Local Similarity 20.3%; Pred. No. 0.059;  
 Matches 86; Conservative 57; Mismatches 156; Indels 124; Gaps 18;

QY 20 VVAHPDKYQVAVLTANG-NIDALYEQCLAHPEYAVVWESKYA-----EPRORI 69  
 Db 50 VVRQGESVEALGLAAGFGDLAGDVGVHAAVAAITGIVDAADADLGAREGGOEV 109  
 QY 70 AASPVADIKVLSGSEALQVATLENVDT--VMAIVGAAGLLPTLLAAKAGKTVLLANKE 127  
 Db 110 AGPGSVDLHVIQLEEGLEHLPDRRLDVRITSAIDHATIDQALVAG-----B 158  
 QY 128 ALVNSGGCIFMGAVSDGAVLPIIDSEHNAIFQCMPPAGTTPGHTAKQARRILLTASGPFR 187  
 Db 159 AIIVE-----QVIAVFAHVV-----RQAVGQFLVQRLGSDHLAGAAGHLL-----GROFR 204  
 QY 188 RPIETLTSSVTPDQAVAH-----PKMDGRKISVDSATMANKGLELIEACLIFN 236  
 Db 205 DOAVEVGTGSHDELRLIALRGHHRAGALDAGR-----ALLVNSAERLHRRRFAE 259  
 QY 237 MEPOIIEV-----VIHPGSI-----HSMVYVDSVLAQWGNPDMRTPIAHAMWPERFDSG 289  
 Db 260 GEVQBMVIAAHVKAATYIVGRHHLAD-----ALGVHDLQIVV--NVAFPQLFLRA 309  
 QY 290 VAPLDIFEVGHMDEKPKDLKFPFCRLAYEAKSGGIMPTVYNANEI-----A 338  
 Db 310 -----QVIRL-----LLGGGHEHPAVLQVALDPTVAGHALADADPA 344

QY 339 VEAPLNEEYK-----TDIAVIERSMAGFPDDA-----GSIEVLQADQD 380  
 Db 345 FEGHIAEQLEFALADGALDHVDVVAVAVDDLAAVAPGSAEADLGSGFYGDLETTILQOEOG 404  
 QY 381 ARE 383  
 Db 405 GGE 407

Search completed: January 29, 2004, 15:57:03  
 Job time: 13.3656 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 26.8953 Seconds  
(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947a-8  
Perfect score: 1985  
Sequence: 1 MKGICILGATSGISVSTLDV.....LQADQDARVARDIKITVA 394

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues  
Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US05\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description       |
|------------|--------|-------------|--------|-------|-------------------|
| 1          | 1985   | 100.0       | 394    | 10    | US-09-934-903-4   |
| 2          | 1985   | 100.0       | 394    | 10    | US-09-934-868-64  |
| 3          | 1985   | 100.0       | 394    | 11    | US-09-941-947a-8  |
| 4          | 1215.5 | 61.2        | 396    | 12    | US-10-381-779-117 |
| 5          | 1052.5 | 53.0        | 394    | 12    | US-10-381-779-119 |
| 6          | 1045   | 52.6        | 398    | 12    | US-10-381-779-100 |
| 7          | 1045   | 52.6        | 398    | 12    | US-10-381-779-118 |
| 8          | 1042   | 52.5        | 405    | 12    | US-10-381-779-122 |
| 9          | 1007   | 50.7        | 397    | 12    | US-10-381-779-99  |
| 10         | 1007   | 50.7        | 397    | 12    | US-10-381-779-120 |
| 11         | 999.5  | 50.4        | 388    | 12    | US-10-381-779-101 |
| 12         | 999.5  | 50.4        | 388    | 12    | US-10-381-779-116 |
| 13         | 994.5  | 45.0        | 394    | 12    | US-10-381-779-97  |
| 14         | 892.5  | 45.0        | 394    | 12    | US-10-381-779-103 |
| 15         | 892.5  | 45.0        | 394    | 12    | US-10-381-779-121 |

|    |       |      |      |    |                     |                    |
|----|-------|------|------|----|---------------------|--------------------|
| 16 | 886   | 44.6 | 402  | 12 | US-10-381-779-102   | Sequence 102, App  |
| 17 | 886   | 44.6 | 402  | 12 | US-10-381-779-123   | Sequence 123, App  |
| 18 | 818.5 | 41.2 | 477  | 9  | US-09-923-556-6     | Sequence 6, Appl   |
| 19 | 818.5 | 41.2 | 477  | 10 | US-09-987-025-2     | Sequence 2, Appl   |
| 20 | 818.5 | 41.2 | 477  | 12 | US-10-381-779-128   | Sequence 128, App  |
| 21 | 817.5 | 41.2 | 477  | 14 | US-10-047-412a-10   | Sequence 10, Appl  |
| 22 | 814.5 | 41.0 | 477  | 9  | US-09-923-556-2     | Sequence 2, Appl   |
| 23 | 813.5 | 41.0 | 477  | 14 | US-10-047-412a-29   | Sequence 29, Appl  |
| 24 | 798.5 | 40.2 | 388  | 12 | US-10-381-779-98    | Sequence 98, Appl  |
| 25 | 798.5 | 40.2 | 388  | 12 | US-10-381-779-125   | Sequence 125, App  |
| 26 | 731.5 | 36.9 | 385  | 12 | US-10-128-713a-4    | Sequence 4, Appl   |
| 27 | 729.5 | 36.8 | 486  | 12 | US-10-381-779-131   | Sequence 131, App  |
| 28 | 721.5 | 36.3 | 487  | 12 | US-10-259-194a-268  | Sequence 268, App  |
| 29 | 718   | 36.2 | 386  | 12 | US-10-381-779-124   | Sequence 124, App  |
| 30 | 697   | 35.1 | 394  | 15 | US-10-156-761-10100 | Sequence 10100, A  |
| 31 | 692   | 34.9 | 392  | 10 | US-09-738-626-5709  | Sequence 5709, App |
| 32 | 691   | 34.8 | 436  | 10 | US-09-712-363-258   | Sequence 258, App  |
| 33 | 691   | 34.8 | 436  | 12 | US-10-381-779-104   | Sequence 104, App  |
| 34 | 691   | 34.8 | 436  | 12 | US-10-381-779-127   | Sequence 127, App  |
| 35 | 681   | 34.3 | 406  | 12 | US-10-381-779-126   | Sequence 126, App  |
| 36 | 620   | 31.2 | 340  | 12 | US-10-289-362-362   | Sequence 362, App  |
| 37 | 596   | 30.0 | 1588 | 12 | US-09-925-778-2     | Sequence 2, Appl   |
| 38 | 486   | 24.5 | 356  | 12 | US-10-381-779-130   | Sequence 130, App  |
| 39 | 473   | 23.8 | 368  | 12 | US-09-882-227-500   | Sequence 500, App  |
| 40 | 428.5 | 21.6 | 177  | 12 | US-10-032-395-1     | Sequence 1, Appl   |
| 41 | 251   | 12.6 | 98   | 12 | US-09-864-408a-8646 | Sequence 8646, App |
| 42 | 114   | 5.7  | 425  | 12 | US-10-369-493-21773 | Sequence 21773, A  |
| 43 | 113.5 | 5.7  | 1073 | 15 | US-10-156-761-12156 | Sequence 12156, A  |
| 44 | 109.5 | 5.5  | 6146 | 15 | US-10-156-761-10436 | Sequence 10436, A  |
| 45 | 109   | 5.5  | 426  | 12 | US-10-369-493-21804 | Sequence 21804, A  |

## ALIGNMENTS

RESULT 1  
US-09-934-903-4  
Sequence 4, Application US/09934903  
Patent No. US20020102690A1  
GENERAL INFORMATION:  
APPLICANT: Koffas, Matthews  
APPLICANT: Odum, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. US20020102690A1on, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Rouviere, Pierre  
APPLICANT: Picotaggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
FILE REFERENCE: C11646 US NA  
CURRENT APPLICATION NUMBER: US/09/934, 903  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229,907  
PRIOR FILING DATE: September 1, 2001  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 4  
LENGTH: 394  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ORF2  
US-09-934-903-4

Query Match 100.0%; Score 1985; DB 10; Length 394;  
Best Local Similarity 100.0%; Pred. No. 7.2e-186;  
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGICILGATSGISVSTLDVARRHPDKYQVAFRANGIDALVYOCCLAHREYVWVMS 60  
DB 1 MKGICILGATSGISVSTLDVARRHPDKYQVAFRANGIDALVYOCCLAHREYVWVMS 60  
QY 61 KVAEFKRIASPVADIKIVLSGEALQOVATLENVDTMAIVGAAGILPTLAARAKAKGT 120

```

Db      61 KVAEFKQRIASPVAADIKVLSSGEALQOYATLENDVTWAAIVGAAGLLPTLAAAKAGKT 120
Qy      121 VILANKELVMSGOIFMOWVSDSGAVLLPISSEHNAIQCPAGYTPGHTAKOARRILLT 180
Db      121 VILANKELVMSGOIFMOWVSDSGAVLLPISSEHNAIQCPAGYTPGHTAKOARRILLT 180
Qy      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVSATMNNKGLIEACLFNNEPD 240
Db      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVSATMNNKGLIEACLFNNEPD 240
Qy      241 QIEVVIHPOSIIHSMVDYDGSVLAQMGKPMRTPIAHAMAPERFDSGVAPLDIFEVGH 300
Db      241 QIEVVIHPOSIIHSMVDYDGSVLAQMGKPMRTPIAHAMAPERFDSGVAPLDIFEVGH 300
Qy      301 MDEFKPDLRFPCLRLAYEAIKSGGIMPTVLANANEIAVEAFINEEVFTDIIVIERSM 360
Db      301 MDEFKPDLRFPCLRLAYEAIKSGGIMPTVLANANEIAVEAFINEEVFTDIIVIERSM 360
Qy      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394
Db      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394

```

## RESULT 2

```

US-09-934-868-64
; Sequence 64, Application US/09934868
; Patent No. US20020137190A1
; GENERAL INFORMATION:
; APPLICANT: Koffas, Matchee
; APPLICANT: Odom, James M.
; APPLICANT: Schenzle, Andreas J.
; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: CL1596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,868
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 64
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Methylomonas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by DXR
US-09-934-868-64

```

Query Match 100.0%; Score 1985; DB 10; Length 394;

Best Local Similarity 100.0%; Pred. No. 7,2e-186;

Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MKGICILGATGSIQVSTLDVVAHHPKDYQVVALTANGNIDALYEQCLAHPEYAVVWMS 60
Db      1 MKGICILGATGSIQVSTLDVVAHHPKDYQVVALTANGNIDALYEQCLAHPEYAVVWMS 60
Qy      61 KVAEFKQRIASPVAADIKVLSSGEALQOYATLENDVTWAAIVGAAGLLPTLAAAKAGKT 120
Db      61 KVAEFKQRIASPVAADIKVLSSGEALQOYATLENDVTWAAIVGAAGLLPTLAAAKAGKT 120
Qy      121 VILANKELVMSGOIFMOWVSDSGAVLLPISSEHNAIQCPAGYTPGHTAKOARRILLT 180
Db      121 VILANKELVMSGOIFMOWVSDSGAVLLPISSEHNAIQCPAGYTPGHTAKOARRILLT 180
Qy      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVSATMNNKGLIEACLFNNEPD 240
Db      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVSATMNNKGLIEACLFNNEPD 240
Qy      241 QIEVVIHPOSIIHSMVDYDGSVLAQMGKPMRTPIAHAMAPERFDSGVAPLDIFEVGH 300
Db      241 QIEVVIHPOSIIHSMVDYDGSVLAQMGKPMRTPIAHAMAPERFDSGVAPLDIFEVGH 300
Qy      301 MDEFKPDLRFPCLRLAYEAIKSGGIMPTVLANANEIAVEAFINEEVFTDIIVIERSM 360
Db      301 MDEFKPDLRFPCLRLAYEAIKSGGIMPTVLANANEIAVEAFINEEVFTDIIVIERSM 360

```

```

Db      301 MDEFKPDLRFPCLRLAYEAIKSGGIMPTVLANANEIAVEAFINEEVFTDIIVIERSM 360
Qy      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394
Db      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394

```

## RESULT 3

```

US-09-941-947a-8
; Sequence 8, Application US/09941947a
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Bzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Matchee
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picatagallo, Steve
; APPLICANT: Riviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947a
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Methylomonas 16a
US-09-941-947a-8

```

Query Match 100.0%; Score 1985; DB 11; Length 394;

Best Local Similarity 100.0%; Pred. No. 7,2e-186;

Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MKGICILGATGSIQVSTLDVVAHHPKDYQVVALTANGNIDALYEQCLAHPEYAVVWMS 60
Db      1 MKGICILGATGSIQVSTLDVVAHHPKDYQVVALTANGNIDALYEQCLAHPEYAVVWMS 60
Qy      61 KVAEFKQRIASPVAADIKVLSSGEALQOYATLENDVTWAAIVGAAGLLPTLAAAKAGKT 120
Db      61 KVAEFKQRIASPVAADIKVLSSGEALQOYATLENDVTWAAIVGAAGLLPTLAAAKAGKT 120
Qy      121 VILANKELVMSGOIFMOWVSDSGAVLLPISSEHNAIQCPAGYTPGHTAKOARRILLT 180
Db      121 VILANKELVMSGOIFMOWVSDSGAVLLPISSEHNAIQCPAGYTPGHTAKOARRILLT 180
Qy      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVSATMNNKGLIEACLFNNEPD 240
Db      181 ASGGPRRTPIETLSSVTPDOAVAHPKMDGKRKISVSATMNNKGLIEACLFNNEPD 240
Qy      241 QIEVVIHPOSIIHSMVDYDGSVLAQMGKPMRTPIAHAMAPERFDSGVAPLDIFEVGH 300
Db      241 QIEVVIHPOSIIHSMVDYDGSVLAQMGKPMRTPIAHAMAPERFDSGVAPLDIFEVGH 300
Qy      301 MDEFKPDLRFPCLRLAYEAIKSGGIMPTVLANANEIAVEAFINEEVFTDIIVIERSM 360
Db      301 MDEFKPDLRFPCLRLAYEAIKSGGIMPTVLANANEIAVEAFINEEVFTDIIVIERSM 360
Qy      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394
Db      361 AQFKPDAGSLELVLOADODAREVARDIITLVA 394

```

## RESULT 4

```

US-10-381-779-117
; Sequence 117, Application US/10381779

```

```

; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessem, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002051
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-381-779-117

```

```

Query Match      61.2%; Score 1215.5; DB 12; Length 396;
Best Local Similarity 64.1%; Pred. No. 2.1e-110;
Matches 245; Conservative 48; Mismatches 88; Indels 1; Gaps 1;

```

```

QY 4 ICIIIGATGSGIVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 63
   7 ISVLAATGSGIGSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 66
   64 EFKORIASPVADIKYVSGSEBALQOVATLENDVTMAAIVGAAGLPTLAAAKGKTYL 123
   67 ALQGLAAGNR-TVIFGBOALCEVASAPEVDVMAAIVGAAGLPTLAAVEAGKTYL 125
   124 ANKEALVMSGOIFMVAQVSDGAVLLPDSSEHNAIFQCPAGYTPGHYAKOARRILLTASG 183
   126 ANKEALVMSGOIFMVAQVSDGAVLLPDSSEHNAIFQCPAGYTPGHYAKOARRILLTASG 185
   184 GPFRTPIETLSSTVTPDOVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 243
   186 GPFRTPIETLSSTVTPDOVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 245
   244 VVHPOSIIHSMVDYVDSVLAQGNPDMPRTPIAHAMAPERFDSGVAPLDIFEVGHMDF 303
   246 VVHPOSIIHSMVDYVDSVLAQGNPDMPRTPIAHAMAPERFDSGVAPLDIFEVGHMDF 305
   304 EKPDKRFPCLRLAYEAIKSGGIMPTVLANANEIVAEFLNEVEFTDIIVIERSSAOF 363
   306 QRPDQRPFCCLRLAYEAIKSGGIMPTVLANANEIVAEFLNEVEFTDIIVIERSSAOF 365
   364 KPDAGSLBLVLAQDQDAREVA 385
   366 AVTAVESLDQVLAADRRARSVA 387

```

```

RESULT 5
US-10-381-779-119
; Sequence 119, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessem, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002051
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119

```

```

; LENGTH: 394
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-381-779-119

```

```

Query Match      53.0%; Score 1052.5; DB 12; Length 394;
Best Local Similarity 55.5%; Pred. No. 2.1e-94;
Matches 216; Conservative 49; Mismatches 123; Indels 1; Gaps 1;

```

```

QY 4 ICIIIGATGSGIVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 63
   6 IELIGSGISIGESTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 65
   64 EFKORIASPVADIKYVSGSEBALQOVATLENDVTMAAIVGAAGLPTLAAAKGKTYL 123
   66 RLEALRRDGT-TVIFGBOALCEVASAPEVDVMAAIVGAAGLPTLAAAKGKTYL 124
   124 ANKEALVMSGOIFMVAQVSDGAVLLPDSSEHNAIFQCPAGYTPGHYAKOARRILLTASG 183
   125 ANKEALVMSGOIFMVAQVSDGAVLLPDSSEHNAIFQCPAGYTPGHYAKOARRILLTASG 184
   184 GPFRTPIETLSSTVTPDOVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 243
   185 GPFRTPIETLSSTVTPDOVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESKVA 244
   244 VVHPOSIIHSMVDYVDSVLAQGNPDMPRTPIAHAMAPERFDSGVAPLDIFEVGHMDF 303
   245 VVHPOSIIHSMVDYVDSVLAQGNPDMPRTPIAHAMAPERFDSGVAPLDIFEVGHMDF 304
   304 EKPDKRFPCLRLAYEAIKSGGIMPTVLANANEIVAEFLNEVEFTDIIVIERSSAOF 363
   305 QRPDQRPFCCLRLAYEAIKSGGIMPTVLANANEIVAEFLNEVEFTDIIVIERSSAOF 364
   364 KPDAGSLBLVLAQDQDAREVA 392
   365 PSDDIGIGGLAQDQDAREVA 393

```

```

RESULT 6
US-10-381-779-100
; Sequence 100, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessem, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002051
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-381-779-100

```

```

Query Match      52.6%; Score 1045; DB 12; Length 398;
Best Local Similarity 54.9%; Pred. No. 1.1e-93;
Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

```

```

QY 1 MKGICIGATGSGIVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVES 60
   1 MKGICIGATGSGIVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVES 60
   61 KVAEFKORIASPVADIKYVSGSEBALQOVATLENDVTMAAIVGAAGLPTLAAAKGKTYL 120
   61 S-ATLKTMLQQQGSFREVLSGGQAAQMAALBVDVDMVAIVGAAGLPTLAAIRAGKT 119

```



```

QY 121 VLIANKALVMSGQITPMQAVSDGAVLLPIDSSEHNAIPQCM--PAGYTPGHTAKQAR--- 175
DB 120 ILLANESLVTGRLPMDAVKOSKQQLPVDSEHNAIPQSLPQPHNIGYADLENGVAV 179
QY 176 RILITASGPFRRPTLETLSVTPDQAVAPKMDGKRISVDSATMMNKGLEIEACLP 235
DB 180 SILITSGGPFRRPTLETLSVTPDQACHPMNSMGRKISVDSATMMNKGLEIEACLP 239
QY 236 NMEPDQIEVVIHPOSIIHSMVDYDGSVLAQMGNPDMRTPIAHAMAMPFRFDSGAVPLDI 295
DB 240 NASASQMEVLIHPOSIVHSWRYQDGSVLAQJGEPDMRTPIAHAMAMPFRVNSGVKPLDF 299
QY 296 FEVGHMDFEKPDLKRPFCLELAYEAIKSGGIMPTVANAANEIVAEFLNEEYKFTDIAYI 355
DB 300 CKLSALTFAAPDYDRPCKLAEAFEQQAATTALANAETIVAAFLAQOIRFTDIAL 359
QY 356 IERSMAQFKPDAGSLVLAQADQDAFEVARDIITKTLVA 394
DB 360 NLSVLEKMDMREPQCVDVLSVDANAREVARKVEMRLAS 398

```

```

RESULT 7
US-10-381-779-118
; Sequence 118, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 118
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-381-779-118

```

```

Query Match 52.6%; Score 1045; DB 12; Length 398;
Best Local Similarity 54.9%; Pred. No. 1.1e-93;
Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

```

```

QY 1 MKGICILGATGSIQVSTLDVVAHNPDKYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 60
DB 1 MKQITLIGSTGICSTLDVVAHNPDKYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 60
QY 61 KVAEFKRIASPVADIKVLSGSEALQVATLENVDYMAAIVGAGLLPTLAAAKGT 120
DB 61 S-AKLKTMLOQGSSTEVLSGQAAQADVALEDVQVAAIVGAGLLPTLAAIRPKGT 119
QY 121 VLIANKALVMSGQITPMQAVSDGAVLLPIDSSEHNAIPQCM--PAGYTPGHTAKQAR--- 175
DB 120 ILLANESLVTGRLPMDAVKOSKQQLPVDSEHNAIPQSLPQPHNIGYADLENGVAV 179
QY 176 RILITASGPFRRPTLETLSVTPDQAVAPKMDGKRISVDSATMMNKGLEIEACLP 235
DB 180 SILITSGGPFRRPTLETLSVTPDQACHPMNSMGRKISVDSATMMNKGLEIEACLP 239
QY 236 NMEPDQIEVVIHPOSIIHSMVDYDGSVLAQMGNPDMRTPIAHAMAMPFRFDSGAVPLDI 295
DB 240 NASASQMEVLIHPOSIVHSWRYQDGSVLAQJGEPDMRTPIAHAMAMPFRVNSGVKPLDF 299
QY 296 FEVGHMDFEKPDLKRPFCLELAYEAIKSGGIMPTVANAANEIVAEFLNEEYKFTDIAYI 355
DB 300 CKLSALTFAAPDYDRPCKLAEAFEQQAATTALANAETIVAAFLAQOIRFTDIAL 359

```

```

QY 356 IERSMAQFKPDAGSLVLAQADQDAFEVARDIITKTLVA 394
DB 360 NLSVLEKMDMREPQCVDVLSVDANAREVARKVEMRLAS 398

```

```

RESULT 8
US-10-381-779-122
; Sequence 122, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 122
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-10-381-779-122

```

```

Query Match 52.5%; Score 1042; DB 12; Length 405;
Best Local Similarity 53.2%; Pred. No. 2.3e-93;
Matches 210; Conservative 68; Mismatches 109; Indels 8; Gaps 3;

```

```

QY 1 MKGICILGATGSIQVSTLDVVAHNPDKYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 60
DB 7 MKQITLIGSTGICSTLDVVAHNPDKYQVVALTANGNIDALYEQCLAHHPREYAVVWMS 66
QY 61 KVAEFKRIASPVADIKVLSGSEALQVATLENVDYMAAIVGAGLLPTLAAAKGT 120
DB 67 AAXMLAEKLNQ-SQTVLVAQQAICELAHHPREYAVVWMSMAIVGAGLLPTLAAAKGT 125
QY 121 VLIANKALVMSGQITPMQAVSDGAVLLPIDSSEHNAIPQCM--PAGYTPGHTAKQAR--- 174
DB 126 VLIANKALVMSGQITPMQAVSDGAVLLPIDSSEHNAIPQCM--PAGYTPGHTAKQAR--- 184
QY 175 RILITASGPFRRPTLETLSVTPDQAVAPKMDGKRISVDSATMMNKGLEIEACLP 234
DB 185 SKIVITSGGPFRRPTLETLSVTPDQACHPMNSMGRKISVDSATMMNKGLEIEACLP 244
QY 235 NMEPDQIEVVIHPOSIIHSMVDYDGSVLAQMGNPDMRTPIAHAMAMPFRFDSGAVPLDI 294
DB 245 FNASAEVLIHPOSIVHSWRYQDGSVLAQJGEPDMRTPIAHAMAMPFRVNSGVKPLDF 304
QY 295 IFVGHMDFEKPDLKRPFCLELAYEAIKSGGIMPTVANAANEIVAEFLNEEYKFTDIAYI 354
DB 305 FYQNLGFTFBPDYQRYCKLADAFSAGYATTAMANAETIVAAFLAQOIRFTDIAL 364
QY 355 IERSMAQFKPDAGSLVLAQADQDAFEVARDIITKTLVA 389
DB 365 LNLQVASTLQPOKIHCHIEDVLEVDKKAERLSQSI 399

```

```

RESULT 9
US-10-381-779-99
; Sequence 99, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1

```

```

CURRENT APPLICATION NUMBER: US/10/381,779
CURRENT FILING DATE: 2003-03-28
PRIOR APPLICATION NUMBER: PCT/US01/30328
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: 60/236,580
PRIOR FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 190
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 99
LENGTH: 397
TYPE: PRN
ORGANISM: Haemophilus influenzae
US-10-381-779-99

```

|                           |        |                  |           |             |
|---------------------------|--------|------------------|-----------|-------------|
| Query Match               | 50.7%; | Score 1007;      | DB 12;    | Length 397; |
| Best Local Similarity     | 50.6%; | Pred. No. 6e-90; |           |             |
| Matches 200; Conservative | 76;    | Mismatches 111;  | Indels 8; | Gaps 3      |

[illegible]

```

RESULT 10
US-10-381-779-120
: Sequence 120: Application US/10381779
: Publication NO. US20030219798A1
: GENERAL INFORMATION:
: APPLICANT: Gokarn, Ravi R
: APPLICANT: Jessem, Holly
: APPLICANT: Zidwick, Mary Jo
: TITLE OF INVENTION: Isoprenoid Production
: FILE REFERENCE: 12904/0020S1
: CURRENT APPLICATION NUMBER: US/10/381,779
: CURRENT FILING DATE: 2003-03-28
: PRIOR APPLICATION NUMBER: PCT/US01/30328
: PRIOR FILING DATE: 2001-09-28
: PRIOR APPLICATION NUMBER: 60/236,580
: PRIOR FILING DATE: 2000-09-29
: NUMBER OF SEQ ID NOS: 190
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 120
: LENGTH: 397
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
US-10-381-779-120

```

| Query Match           | Score     | IDB    | Length |
|-----------------------|-----------|--------|--------|
| Best Local Similarity | Pred. No. | 6e-90; |        |
| 50.78;                | 1007;     | DB 12; | 397;   |
| 50.68;                |           |        |        |

|    | Matches | 200, | Conservative  | 76, | Mismatches | 111, | Indels | 8, | Gaps | 3, |
|----|---------|------|---|-----|------------|------|--------|----|------|----|
| QY |         | 2    | KGICILANTSGISGISTUDVYARABDKQVVALTNANGIDALTPOCLAHHEBYAVNMESK     |     | 61         |      |        |    |      |    |
| DB |         | 4    | QNIYILASTSGISKSTLSVLENPFQXTHAPLVAGKNVEMPECKIFRHFPAALDDVNA       |     | 63         |      |        |    |      |    |
| QY |         | 62   | VAEYKORILASPVDIVYLSGSBALQGVATLENDVTMAALVCAAGLLPTLAAKAGTV        |     | 121        |      |        |    |      |    |
| DB |         | 64   | AKILRECLIAHHT-PTLEVLAGRALCELAHPDDQIMASTVGAAGLLPTLSAYKAKRV       |     | 122        |      |        |    |      |    |
| QY |         | 122  | LLANKREALVMSGOIFMQAVSDSGAVILLPIDSENALFOCMP-----AGYTPGHTAQAR     |     | 175        |      |        |    |      |    |
| DB |         | 123  | LLANKREALVTCQGFIDAVKNYGSKLLPYDSENHALFOSTPEPEAQKIGFCP-LSELGVS    |     | 181        |      |        |    |      |    |
| QY |         | 176  | RIILTAGSGPPRRPRITPLTSSVDPDQVAPRKMGKISVDSATMNNKGLIELACILF        |     | 235        |      |        |    |      |    |
| DB |         | 182  | KIILTSGSGPPRYPLTFOFNITLPEQVVAHPHNSMGKISVDSATMNNKGLIYEAKRLF      |     | 241        |      |        |    |      |    |
| QY |         | 236  | NMEPDOILEVVHPQSILHSMWDYVDSVLAAQWGNPDKRTPLAHAMAPERFDSGVAPDI      |     | 295        |      |        |    |      |    |
| DB |         | 242  | NASAEEMEVIIHPQSILHSMRVYDGSVLIQMGNPDKRTPLAHAMAPHRPAGVEPDLF       |     | 301        |      |        |    |      |    |
| QY |         | 296  | FEVGHDPFEKPDLRFPOLRLAYEALTSGGIMPTVLNAAEIAVEATLNEBVKETDLAVI      |     | 355        |      |        |    |      |    |
| DB |         | 302  | FKIETELPIEDFPFRYTNLKATIDAFAPAGOVATLTANNAANAEIAVQAFLDRIQGIWDIAKI |     | 361        |      |        |    |      |    |
| QY |         | 356  | IERMAQPKPPDDAGSLTELVQAQDDAPARVADITK                             |     | 390        |      |        |    |      |    |
| DB |         | 362  | NSKTIERSPTTIONIDVLEIDQRBBIATKLTKLR                              |     | 396        |      |        |    |      |    |

```

US-10-381-779-101
RESULT 11
Sequence 101, Application US/10381779
Publication No. US20030219798A1
GENERAL INFORMATION:
APPLICANT: Gokhetti, Ravi R
APPLICANT: Jessen, Holly
APPLICANT: Zidwick, Mary Jo
TITLE OF INVENTION: Isoprenoid Production
FILE REFERENCE: 12904/0020US1
CURRENT APPLICATION NUMBER: US/10/381,779
CURRENT FILING DATE: 2003-03-28
PRIOR APPLICATION NUMBER: PCT/US01/30328
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: 60/736,580
PRIOR FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 190
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 101
LENGTH: 386
TYPE: PRT
ORGANISM: Zymonas mobilis
US-10-381-779-101

```

|                       |                 |                   |          |            |
|-----------------------|-----------------|-------------------|----------|------------|
| Query Match           | 50.4%           | Score 999.5       | DB 12    | Length 388 |
| Best Local Similarity | 50.9%           | Pred. No. 3,26-89 |          |            |
| Matches 199           | Conservative 74 | Mismatches 109    | Indels 9 | Gaps 3     |

  

|    |     |  |     |
|----|-----|--|-----|
| QY | 4   | ICIIAGTSGTGVSTLDVYARHPDKYQVVALTANGNIDALTYQCCLAHPEBYAVVYMSKVA | 63  |
| DB | 7   | VTVIAGTSGTSHSTLIDLERNLDRQVVALTANNVVKDLDAAKRTNKKRVILADPSLYN   | 66  |
| QY | 64  | EFKQRIASPVAIDIKVLSGSEALDOVLTLENVDPVMAIVGAGLLPTLTAARAKGVTL    | 123 |
| DB | 67  | DLKEALGSG---SVEAARAGADALVEAAMM--GADVTMAIITCAGIKATLAIIRKKTVAL | 122 |
| QY | 124 | ANKKALVWSGQIFQMAVDSGAVLLPFDISENALFOCPGAGTGGHTKQARRILLTASG    | 183 |
| DB | 123 | ANKSLVWSAGLMDAIVREHGTLLLPDPSHNALFOCF-----PHNRDYYRRIITITNSG   | 177 |
| QY | 184 | GPFRRTPLTTLSSVTPDQAVAHFPMNDGKRLSVGSATIMAKGLILIECLFNNEDDIE    | 243 |

[illegible]

```

RESULT 12
US-10-381-779-116
; Sequence 116, Application US/10381779
; Publication No. US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gohain, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/50328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 116
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Zymomonas mobilis
; US-10-381-779-116

```

|    | Query Match           | 50.4%;                 | Score 999.5;   | DB 12;    | Length 388; |
|----|-----------------------|------------------------|--|-----------|-------------|
|    | Best Local Similarity | 50.9%;                 | Pred. 3.2e-8;  |           |             |
|    | Matches 199;          | Conservative 74;       | Mismatches 109;  | Indels 9; | Gaps 3      |
| QY | 4                     | ICILGATGSI             | GVSTLDVAAKPPDCKTQVVALPANKGIDALYACQLAHPELVAVVMESKVA     | 63        |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
|    | 7                     | VYVLGATGSI             | GHSTLDLIERLNDRYQVATLANKRVYDLDAAKRTAKAVAVINDPSLYN       | 66        |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| QY | 64                    | EFKQRTAA               | SPVADIKVLSGSEALQOAVTLLENDVTMAAIVGAAGLPLTAAAKGKTVLL     | 123       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| Db | 67                    | DLKEALAGS---           | SVEAAAGADLVAAAM--GADTMAAIIICAGAGKATLAIKRGKTVAL         | 122       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| QY | 124                   | ANKAELWMSGOIFMCAVPSDGA | VLLPIDSEHNAIFQCMAGYTPGHTAKQARRILLTASG                  | 183       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| Db | 123                   | ANKESIVSAGGLMDIAVR     | EHGTTLLPVDSEHNAIFQC-----PHHNRDYYRRIITIASG              | 177       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| QY | 184                   | GPRRTPI                | LTSSVTPDDVAHAPKMDNGRKISVDSATPAKGLBLIACLLFNKPEQITE      | 243       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| Db | 178                   | GPRRTISLAEMKIVT        | PERAVQHPNMMSGAKSISDSATPMKGLBLIEAYHFOIPEKFE             | 237       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| QY | 244                   | VTHIPOSII              | HSWVDYDGSVLAQMGNPMPKPTIADAMAMPKPSGVAALDIFEGVGMDF       | 303       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| Db | 238                   | LVHPQSV                | HSWVEYHDSILMQISPDKRTPIGHTLAMPKMETPRAESIDFTKLKMDF       | 297       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| QY | 304                   | EKEDLR                 | RFPLRLAIEAIIKSGGIMPTVLNAANEIYAENFINEEYKFTDIAVILERSNAQF | 363       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| Db | 298                   | BAEDYER                | FALTIAMESIKSGARPAVMMAANEIYAFAFDKTIIGFDIAKIAYEKTIDHY    | 357       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| QY | 364                   | KPDAGSL                | ELVYLAQODDAEVARDIKTLVA                                 | 394       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |
| Db | 358                   | TPATPSS                | LEDVPAIDNEARIQAALAMESIPAA                              | 388       |             |
|    |                       | : : : : : :            | : : : : : :  |           |             |

```

US-10-381-779-97
: Sequence 97, Application US/10381779
: Publication No. US20030219798A1
: GENERAL INFORMATION:
: APPLICANT: Gokarrn, Ravi R
: APPLICANT: Jessen, Holly
: APPLICANT: Zidwick, Mary Jo
: TITLE OF INVENTION: Isoprenoid Production
: FILE REFERENCE: 12904/002051
: CURRENT APPLICATION NUMBER: US/10/381,779
: CURRENT FILING DATE: 2003-03-28
: PRIOR APPLICATION NUMBER: PCT/US01/30328
: PRIOR FILING DATE: 2001-09-28
: PRIOR APPLICATION NUMBER: 60/236,560
: PRIOR FILING DATE: 2000-09-29
: NUMBER OF SEQ ID NOS: 190
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 97
: LENGTH: 386
: TYPE: PRF
: ORGANISM: Sphingomonas trueperi
US-10-381-779-97

```

```

Query Match 50.1%; Score 994.5; DB 12; Length 386;
Best Local Similarity 51.0%; Pred. No. 9.7e-89;
Matches 203; Conservative 69; Mismatches 109; Indels 17; Gaps 4

QY      1 MKGICILGATSGIGSVTLDVVARHPDKYQVVALTANGNTIDALYEQCLAHFEIYAVYKES 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2 VKRVTYVIGATGSGVGTSLTDLIERNPBAFEVVALTANGCIVEKILAAAIITRARCAYVADEK 61
QY      61 KVAEFGKRIASPVADIKYLSGSALQGVATLENVDVTMAIVGAGLPLTAAKAGKT 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62 CLPALGELLGSG--GVAWGGAHSVCDVARN-GADWTMAIVSGAGKPYMAALBRAGT 117
QY      121 VLTANKEALVMSGQIFMQAVSDSGAVLLPIDSEHNAIFQCM----PAGYTPKHTAKQARR 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      118 VALANKESIVSAGEVWMAAARAHGATILLPVDSERNAVFQCIDRTAPRG-----VRR 168
QY      177 ILLTASGSPRRPTIETLSYTPPOVAHPHMDGKRI SVNSATMMKGLIECLLEN 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      169 ILLTASGSPRRPTPKEMARDITTPQAQAAHPWMSGAKI SVNSATMMKGLIEFHLFP 228
QY      237 MEPCQIEVAVIHPOSIHSNVDYVNGSVLAQGNPDMPETPIAHAMAPERFDSGVALPIDF 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      229 VAAEQLVAVLTHRQSVVMSMEYVDGSLAQGTGTDMPETPIAYALAMPERMETLCPDLDA 288
QY      297 EYGMDFEKDLPKFPCLRLAYEALKEGGIMFTVLANAEIAYEAFLEBVKFTDIAYII 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      289 TVGGLFEENFDLPFPALALAMEALKKAGARFALINANEVAFAVAFLAGRIGLEIAAIS 348
QY      357 ERSMAQKFPDDAGSLSEVLQADQDARVARDIITIVA 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      349 ADTLSTRYPAPETLDAVLAIDAERLYAARVDCVA 386

RESULT 14
US-10-381-779-103
: Sequence 103, Application US/10381779
: Publication No. US20030215798A1
: GENERAL INFORMATION:
: APPLICANT: Gokarm, Ravi R
: APPLICANT: Joessem, Holly
: APPLICANT: Zidwick, Mary Jo
: TITLE OF INVENTION: Isoprenoid Production
: FILE REFERENCE: 12904/002US1
: CURRENT APPLICATION NUMBER: US/10/381,779
: PRIOR FILING DATE: 2003-03-28
: PRIOR APPLICATION NUMBER: PCT/US01/30328
: PRIOR FILING DATE: 2001-09-28
: PRIOR APPLICATION NUMBER: 60/236,560
: PRIOR FILING DATE: 2000-09-29
: NUMBER OF SEQ ID NOS: 190

```

Thu Jan 29 17:41:30 2004

us-09-941-947a-8.rapht

Page 7

```

: SOFTWARE: FASTSHQ for Windows Version 4.0
: SEQ ID NO: 103
: LENGTH: 394
: TYPE: CRT
: ORGANISM: Synechocystis sp. PCC 6803
US-10-381-779-103

```

|                       |                  |                  |            |             |
|-----------------------|------------------|------------------|------------|-------------|
| Query Match           | 45.0%;           | Score 892.5;     | DB 12;     | Length 394; |
| Best Local Similarity | 47.3%;           | Pred. No. 1e-78; |            |             |
| Matches 183;          | Conservative 63; | Mismatches 128;  | Indels 13; | Gaps 3;     |

```

QY      1 MKKICILGAGTSGISGATLLVVAHEDPKQVAVALTANRIDLAYECIAHPBEVAVVMES 60
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      2 VKRISILSGTSGISGCTGLDIIIVTHHPDAFQVGLAAGGVALLAQVAFREIRVAINQAE 61
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61 KVAEFFQRIAPASPVADIKVLSSGEALQVATLENDVTMAAIVGAGILPTLAAAKGKT 120
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      62 KLEEDLKAAAVELFDYQPMKYVEEGVEVFAARXGDASVVTGIVGAGILPTLMAAIAAGKD 121
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121 VLLANKAALMSQIINQAVSOSGAVILPISDENHAISQCM---PAGTTPCHTKQAKRI 177
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      122 IALANKETLIIAGAVVPLPVEKMGVCLPADSESHAIPOCCLQGVPEG-----GIKRI 173
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      178 LITASGSPERRPIETLSGSVTPDOAAVAHPKWMGKRIISVDSATMMNKLELIEACILPNN 237
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      174 ILTPASGARFDLPVELLPVTVQDALKHPNMMSGKTIIDSITLANKELVTEAHYLPGL 233
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      238 EPDQIEVVIHPQSIHSMVDYVDSVLAQMGNDPKRTEIAHAMWPERFDSGVAPLIDIFE 297
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      234 DYDHIDIVIHPOSIHSLIEVQDTSVLAQMGPDWRLLPILVATSPERIVYDMGEPDLVYK 293
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      298 VGMADDEKBPDLKRFPLCLRLAYELIASSGIMPTVLAANEIAVEAFIAEVEAFVTDIANTIE 357
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      294 AGSISFREBPDDPKYPCMLAYAGGRAGGAMPVILANAEQAVALFLORKISFLDIPRLIE 353
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      358 RSMAGFPPDAGS--LELVLAQADQADR 382
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      354 KTCDLIVGQNTASPDLEITLLAADQAKR 380
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

1  RESULT 15
2  US-10-381-779-121
3  Sequence 121, Application US/10381779
4  Publication No. US20030219798A1
5
6  GENERAL INFORMATION:
7
8  APPLICANT: Gokarn, Havi R
9  APPLICANT: Jessen, Rolly
10 APPLICANT: Zidwick, Mary Jo
11 TITLE OF INVENTION: Isoprenoid Production
12 FILE REFERENCE: 12904/002US1
13 CURRENT APPLICATION NUMBER: US/10/381, 779
14 CURRENT FILING DATE: 2003-03-28
15 PRIOR APPLICATION NUMBER: PCT/US01/30328
16 PRIOR FILING DATE: 2001-09-28
17 PRIOR APPLICATION NUMBER: 60/236,580
18 PRIOR FILING DATE: 2000-09-29
19 NUMBER OF SEQ ID NOS: 190
20 SOFTWARE: FastSeq for Windows Version 4.0
21 SEQ ID NO 121
22 LENGTH: 394
23
24 TYPE: RNT
25 ORGANISM: Synechocystis sp. PCC 6803
26 US-10-381-779-121

```

|                       |                  |                  |            |             |
|-----------------------|------------------|------------------|------------|-------------|
| Query Match           | 45.0%;           | Score 892.5;     | DB 12;     | Length 394; |
| Best Local Similarity | 47.3%;           | Pred. No. 1e-78; |            |             |
| Matches 183;          | Conservative 63; | Mismatches 128;  | Indels 13; | Gaps 3      |

```
QY      1 MKGICLIGATGSGVSTLDVVAHPDKYGVALLTANGNIDALYECCAAHBEYVVMMS 60
      2 VKRISITGSGTSGIGTQTLDTVTHHPAPQYGVGLAGNALLAQOVAEFREIYAIQAE 61
QY      61 KVAEFKQRIAAEVALIKVLSSSEALQQVATLENDPTMAALVGAAGLLPTLAARKAGT 120
```

```

Db      62  K4EDLKAAVAELTDYQPMVYVGBSEBVEVAREXGAEVSVTGIVGQALPFLMATAIAGKD 121
Qy      121  VLLANKEALWMSGQIFMQAVSDSGAVLLPISEHNAIFQCM---PAGTTPHTKQARRI 177
Db      122  IATANRETLIAGAPVYLLPIVERMGYKLLPAOSEHSAIFQCLQGVPEG-----GURRI 173
Qy      178  ILTASGAPERRATPIETLSVTPDQAVAPKPDMDGKISVDATMMNKLELIEACILFNM 233
Db      174  ILTASGAFRLDPEVRLPFVYQDLAKHPNMMSGOKIITDATALNKGLEVEIAHYIFGL 233
Qy      238  EPDQIEVVIHPOSIITHSMVDYVDSVLAQMGNPDKRTPIAHMAMPERRFDSGVAPLDIPE 297
Db      234  DYDHIIIVIHPOSIITHSLIEVQDTSVLALQGLPMDRLEPLALALWPERIYIDMDPLDVK 293
Qy      298  VGHMDREKEDLKRFPCLRLATAYALATSGGIMPTVANAELIVEKFLINEVVFUJAVIE 355
Db      294  AGSELSERPEDHDKYRPMQIAYGAGAGAGAMPAYVANAENQAVALLFLOKJSLFDIPLRIE 355
Qy      358  RSMAGKFPDDAGS--LELVQADQDAR 382
Db      354  KTCDLYVGQNTASPDLETLIANDQMAR 380

```

Search completed: January 29, 2004, 16:21:15  
Job time : 27.8953 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 11.5928 Seconds  
(without alignments)  
3268.453 Million cell updates/sec

Title: US-09-941-947a-8  
Perfect score: 1985  
Sequence: 1 MGICICLGATGSGIGVSTLDV.....LQADQAREVARDIKITLVA 394

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID    | Description        |
|------------|--------|-------------|--------|----------|--------------------|
| 1          | 1215.5 | 61.2        | 396    | 2 E83188 | 1-deoxy-D-xyloose  |
| 2          | 1052.5 | 53.0        | 394    | 2 B82000 | 1-deoxy-D-xyloose  |
| 3          | 1045   | 52.6        | 398    | 2 B64741 | YaeM protein - Bac |
| 4          | 1044   | 52.6        | 398    | 2 G90501 | 1-deoxy-D-xyloose  |
| 5          | 1044   | 52.6        | 398    | 2 G85501 | 1-deoxy-D-xyloose  |
| 6          | 1040.5 | 52.4        | 394    | 2 A81229 | 1-deoxy-D-xyloose  |
| 7          | 1027   | 51.7        | 398    | 2 AF0529 | 1-deoxy-D-xyloose  |
| 8          | 1023   | 51.5        | 402    | 2 D82099 | 1-deoxy-D-xyloose  |
| 9          | 1020   | 51.4        | 398    | 2 AG0128 | 1-deoxy-D-xyloose  |
| 10         | 1007   | 50.7        | 397    | 2 A64014 | conserved hypotet  |
| 11         | 9941   | 47.4        | 396    | 2 H87228 | 1-deoxy-D-xyloose  |
| 12         | 892.5  | 45.0        | 394    | 2 S76331 | hypothetical prote |
| 13         | 891    | 44.9        | 398    | 2 F84957 | 1-deoxy-D-xyloose  |
| 14         | 865    | 43.6        | 399    | 2 E82349 | deoxyxyloose 5-ph  |
| 15         | 857.5  | 43.2        | 399    | 2 H87486 | 1-deoxy-D-xyloose  |
| 16         | 854    | 43.0        | 395    | 2 E97121 | 1-deoxy-D-xyloose  |
| 17         | 849    | 42.8        | 397    | 2 AC2897 | 1-deoxy-D-xyloose  |
| 18         | 849    | 42.8        | 397    | 2 E97572 | 1-deoxy-D-xyloose  |
| 19         | 818.5  | 41.2        | 406    | 2 T52570 | 1-deoxy-d-xyloose  |
| 20         | 816.5  | 41.1        | 380    | 2 A81239 | deoxyxyloose 5-ph  |
| 21         | 812.5  | 40.9        | 380    | 2 A11601 | deoxyxyloose 5-ph  |
| 22         | 800    | 40.3        | 365    | 2 E83522 | 1-deoxy-d-xyloose  |
| 23         | 798.5  | 40.2        | 388    | 2 B69881 | conserved hypotet  |
| 24         | 709.5  | 35.7        | 379    | 2 C81578 | 1-deoxy-D-xyloose  |
| 25         | 707.5  | 35.6        | 379    | 2 B72091 | CT071 hypothetical |
| 26         | 706.5  | 35.6        | 379    | 2 G86533 | 1-deoxy-D-xyloose  |
| 27         | 698    | 35.2        | 396    | 2 F75388 | hypothetical prote |
| 28         | 691    | 34.8        | 396    | 2 A70923 | 1-deoxy-D-xyloose  |
| 29         | 684.5  | 34.5        | 379    | 2 G81712 | 1-deoxy-D-xyloose  |

|       |      |      |   |        |                    |
|-------|------|------|---|--------|--------------------|
| 681   | 34.3 | 406  | 2 | A87107 | hypothetical prote |
| 646.5 | 32.6 | 379  | 2 | A71562 | hypothetical prote |
| 641   | 32.3 | 380  | 2 | F70336 | conserved hypotet  |
| 629   | 31.7 | 383  | 2 | C71304 | conserved hypotet  |
| 547   | 27.6 | 376  | 2 | B72321 | conserved hypotet  |
| 486   | 24.5 | 356  | 2 | B81278 | probable 1-deoxy-D |
| 473   | 23.8 | 368  | 2 | H64546 | conserved hypotet  |
| 455   | 22.9 | 368  | 2 | G71961 | hypothetical prote |
| 133.5 | 6.7  | 443  | 2 | S03257 | methyl coenzyme M  |
| 114   | 5.7  | 522  | 2 | C82446 | probable gamma-glu |
| 111   | 5.6  | 3104 | 2 | S20473 | fatty-acid synthas |
| 110.5 | 5.6  | 367  | 2 | AD1687 | prephenate dehydro |
| 108   | 5.4  | 296  | 2 | AF2902 | conserved hypotet  |
| 106   | 5.3  | 456  | 2 | G97677 | hypothetical prote |
| 105.5 | 5.3  | 923  | 2 | S44226 | periodic cryptopha |
|       |      | 367  | 2 | AD1315 | prephenate dehydro |

ALIGNMENTS

|       |  |   |
|-------|--|---|
| ES31  | 1  |   |
| 1-dec | y-d-xylose 5-phosphate reductoisomerase PA3650 [imported] - Pseudomonas aeruginos    |   |
| C:Spa | ies: Pseudomonas aeruginosa  |   |
| C:Dat | : 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000                |   |
| C:Acc | sion: E83188   |   |
| R:Stc | er: C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brj |   |
| atman | S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,    |   |
| .; Lk | Y.; S.; Olson, M.V.  |   |
| Natut | 406, 959-964, 2000   |   |
| A:Tit | e: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathoc  |   |
| A:Ref | rence number: A82950; PMID:20437337; PMID:10984043                                   |   |
| A:Acc | sion: E83188   |   |
| A:Stc | us: preliminary  |   |
| A:Mo  | cule type: DNA   |   |
| A:Ref | dues: 1-396 <STO>  |   |
| A:Cre | s-references: GB:AE004785; GB:AE004091; NID:99949809; PIDN:AA07038.1; GSPDB:GN001    |   |
| A:Exi | perimental source: strain PA01   |   |
| C:Gen | tics:  |   |
| A:Gen | : dxt; PA3650  |   |
| C:Seq | family: conserved hypothetical protein H1807   |   |
| Que   | y Match  | 61.2%; Score 1215.5; DB 2; Length 396;                            |
| Be:   | Local Similarity   | 64.1%; Pred. No. 4.9e-84;   |
| Ma:   | hes 245; Conservative  | 48; Mismatches 88; Indels 1; Gaps 1;                              |
| Qy    | 4  | ICILGANTSGIGVSTLDVYARHPDKQVVALFANGNIDALYEOCLAHHPYAVVMESKVA 63     |
| Db    | 7  | ISVLGATSGIGVSTLDVYVRHPDRYKAFALNFGSLAEALCLRRRPYAVVPEOAAAI 66       |
| Qy    | 64   | EKKRIASPVADIRVLSGSEALQCVATLENDVTMAAIVGAAGLLPTLAARAKGTVLL 123      |
| Db    | 67   | ALQGSILAAAGR-TRVLFGEOALCEVASAPEVDYMAAIVGAAGLPSTLAAEAGRVLL 125     |
| Qy    | 124  | AKKEALVMSGQIFPMQAVDSDGAVLLPISSEHNAIFQCPAGTTPGHTAKQARRILLTASG 183  |
| Db    | 126  | AKKEALVMSGALLFPMQAVRSGAVLLPISSEHNAIFQSLPRYVADGDERGVARRILLTASG 185 |
| Qy    | 184  | GPFRPTPIETLTSVTPDQAVAHPPKMDMKRKISVDSATVMNKGLELIEACLPFMNEPDQIE 243 |
| Db    | 186  | GPFRPTPIEQQLASVTFEQACAHPPNMSMRKISVDSASNNKLELIEACLPFMNEPDQIE 245   |
| Qy    | 244  | VTHPQSIHSMVDVVDGSVLAQKGNPMKPTPIAAMAMPFSDGVAFLDIEFGHMF 303         |
| Db    | 246  | VTHPQSIHSMVDVVDGSVLAQKGNPMKPTPIAAMAMPFSDGVAFLDIEFGHMF 305         |
| Qy    | 304  | EKKDKRFFCLLAEYAIKSGIMPTVLAANEIYVEAFINEEVKFTDIAVLIERSMAQF 363      |
| Db    | 306  | QRPDORFFCLLAEYAIKSGIMPTVLAANEIYVEAFINEEVKFTDIAVLIERSMAQF 365      |
| Qy    | 364  | KPDNAGSELEVTLQADQAREVA 385  |

Db 366 AWTAVESIDQVLAADRARSVA 367

# RESULT 2

882000

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: B82000

R:Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

R:Holroyd, S.; Jagsall, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; M01D:2022256; PMID:10761919

A:Accession: B82000

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <PAR>

A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83399.1; PID:g737885

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: dxt; NMA0083

C:Superfamily: conserved hypothetical protein HI0807

C:Keywords: oxidoreductase

Query Match 53.0%; Score 1052.5; DB 2; Length 394;

Best Local Similarity 55.5%; Pred. No. 9.5e-72;

Matches 216; Conservative 49; Mismatches 123; Indels 1; Gaps 1;

4 ICLGATGSGVSTLDVVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNESKVA 63

6 LTLTSTGSGVSTLDVVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNESKVA 65

64 EKQRIASPVADIKYLSSEALQVATLANTDYMAALVGAAGLLPTLAARKAKTYLL 123

66 RLEALKRDGTA-TYVTLHAQALVDVASADEVSGWCAVAGVLPALAAQKRTYLL 124

124 AKKALVMSGQIFMOAVSDGAVLLPIDEHNALFQCPDAGTTPGHTAKQARRILLTASG 183

125 AKKALVMSGQIFMOAVSDGAVLLPIDEHNALFQCPDAGTTPGHTAKQARRILLTASG 184

184 GFFRTPIETLSVTPDQAVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNESKVA 243

185 GFFRTPIETLSVTPDQAVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNESKVA 244

244 VVIHPSIHSWVDYDGSVLAQMGNPKRTPIAHAMAPERFDGVAFLDIFEGHDF 303

245 VVIHPSIHSWVDYDGSVLAQMGNPKRTPIAHAMAPERFDGVAFLDIFEGHDF 304

304 EKQRIASPVADIKYLSSEALQVATLANTDYMAALVGAAGLLPTLAARKAKTYLL 363

305 EKQRIASPVADIKYLSSEALQVATLANTDYMAALVGAAGLLPTLAARKAKTYLL 364

364 EKQRIASPVADIKYLSSEALQVATLANTDYMAALVGAAGLLPTLAARKAKTYLL 392

365 FSDGIGDIGGLAQDARTAAQAARFTGL 393

# RESULT 3

B64741

C:Species: Escherichia coli (strain K-12)

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: B64741

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: B64741

A:Accession: B64741

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-398 <BLAT>

A:Cross-references: GB:AE000126; GB:U00096; NID:g1786358; PIDN:AACT3284.1; PID:g1786369;

A:Experimental source: strain K-12, substrain MGI655

C:Genetics:

A:Gene: yaem

C:Superfamily: conserved hypothetical protein HI0807

C:Keywords: nucleotide binding; P-loop

F11-119/Region: nucleotide-binding motif A (P-loop)

Query Match 52.6%; Score 1045; DB 2; Length 398;

Best Local Similarity 54.9%; Pred. No. 3.5e-71;

Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;

1 MKGICILGATGSGVSTLDVVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNES 60

1 MKGICILGATGSGVSTLDVVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNES 60

61 KVAEFKQRIASPVADIKYLSSEALQVATLANTDYMAALVGAAGLLPTLAARKAKGT 120

61 S-AKLTKTMLQGGQSRREVLSGGQACDMALBEDVDQVMAIYGAAGLLPTLAIRAGKT 119

121 VLANKALVMSGQIFMOAVSDGAVLLPIDEHNALFQCPDAGTTPGHTAKQARRILLTASG 175

120 ILANKALVMSGQIFMOAVSDGAVLLPIDEHNALFQCPDAGTTPGHTAKQARRILLTASG 179

176 RILTTAGCGFFRTPIETLSVTPDQAVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNES 235

180 SILTTAGCGFFRTPIETLSVTPDQAVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNES 239

236 MNEPDQIEVTHPSITHSWVDYDGSVLAQMGNPKRTPIAHAMAPERFDGVAFLDIFEGHDF 295

240 NASASQMEVTHPSITHSWVDYDGSVLAQMGNPKRTPIAHAMAPERFDGVAFLDIFEGHDF 299

296 FEVGHMDFKRPDCLRLAYEAIKSGGIMPTLVANAEIYEAFLNEEYKFTDIYV 355

300 CKSLATFPADYRYCYCLKLAEMAFEGQALTTALNAAEITFAALQCIIRTDIAL 359

356 IERSMAQFDDAGSLVLAQDDARFVARDIKITLV 394

360 NLVLEKMDKEEPQCVDDVLSVDANAREVARKVEYRLAS 398

RES T 4

G300

1-d xy-D-xylose 5-phosphate reductoisomerase (imported) - Escherichia coli (strain O1

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: G90650

R:Hashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gata, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

es, 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; M01D:21156231; PMID:11258796

A:Accession: G90650

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-398 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA333598.1; PID:g13359631; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0505952

C:Genetics:

A:Gene: ECs0175

C:Superfamily: conserved hypothetical protein HI0807

Query Match 52.6%; Score 1044; DB 2; Length 398;

Best Local Similarity 54.9%; Pred. No. 4.2e-71;

Matches 219; Conservative 55; Mismatches 119; Indels 6; Gaps 3;

1 MKGICILGATGSGVSTLDVVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNES 60

1 MKGICILGATGSGVSTLDVVARHPDKYQVAALFRANGNIDALVEQCLAHPEYAVVNES 60

61 KVAEFKQRIASPVADIKYLSSEALQVATLANTDYMAALVGAAGLLPTLAARKAKGT 120

61 S-AKLTKTMLQGGQSRREVLSGGQACDMALBEDVDQVMAIYGAAGLLPTLAIRAGKT 119

QY 121 VLLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 175  
 DB 120 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 179  
 QY 176 RLLTASGGPFRRTPIETLSVTPDQAVAHFPMKDMGRKISVDSATMMNKGLELLACLLF 235  
 DB 180 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 239  
 QY 236 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 295  
 DB 240 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 299  
 QY 296 FEVGHMDEKPPDLKPPCRLAYEALIKSGGIMPTVLNANEIAYEAFLEBEVKFTDIYVI 355  
 DB 300 CRLSALTFAPPYDRYPCIKLAMEAFEGQQAATTLNANETLVAALAQIRFTDIAL 359  
 QY 356 IERSNAQFPDAGSLBLVLAQDQAREVARDIITLVA 394  
 DB 360 NLVLEKMDKREPCQVDVLSVDASAREVARKVRLAS 398

RESULT 5  
 G85501  
 1-deoxy-D-xylulose 5-phosphate reductoisomerase [similarity] - Escherichia coli (strain  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Nov-2001  
 C:Accession: G85501  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Gotohbeck, E.J.; Davis, N.W.; Lin, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: AB5480; MUID:21074935; PMID:11206551  
 A:Accession: G85501  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-398 <STO>  
 A:Cross-references: GB:AB005174; NID:912512898; PIDN:AA654475.1; GSPDB:GN00145; UMG:Z01  
 A:Experimental source: strain O157:H7, substrain ED1933  
 C:Genetics:  
 A:Gene: yaeM  
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 52.6%; Score 1044; DB 2; Length 398;  
 Best local similarity 54.9%; Pred. No. 4, 2e-71;  
 Matches 219; Conservative 55; Mismatches 119; Indels 6; Gaps 3;  
 QY 1 MGICIGATGSGITLVDVARRHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMS 60  
 DB 1 MGICIGATGSGITLVDVARRHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMS 60  
 QY 61 KVAEFKRIASPVADIKVLGSGEALQVATLENDVYMAIVGAAGLLPTLAAKAGKT 120  
 DB 61 KVAEFKRIASPVADIKVLGSGEALQVATLENDVYMAIVGAAGLLPTLAAKAGKT 120  
 QY 61 S-AKLKTLVLOOGSGRTVLSGOAACAADMALEVDQYMAAIVGAAGLLPTLAAKAGKT 119  
 DB 61 S-AKLKTLVLOOGSGRTVLSGOAACAADMALEVDQYMAAIVGAAGLLPTLAAKAGKT 119  
 QY 121 VLLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 175  
 DB 120 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 179  
 QY 176 RLLTASGGPFRRTPIETLSVTPDQAVAHFPMKDMGRKISVDSATMMNKGLELLACLLF 235  
 DB 180 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 239  
 QY 236 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 295  
 DB 240 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 299  
 QY 296 FEVGHMDEKPPDLKPPCRLAYEALIKSGGIMPTVLNANEIAYEAFLEBEVKFTDIYVI 355  
 DB 300 CRLSALTFAPPYDRYPCIKLAMEAFEGQQAATTLNANETLVAALAQIRFTDIAL 359  
 QY 356 IERSNAQFPDAGSLBLVLAQDQAREVARDIITLVA 394

DB 360 NLVLEKMDKREPCQVDVLSVDASAREVARKVRLAS 398  
 RESULT 6  
 A8122  
 1-deoxy-D-xylulose 5-phosphate reductoisomerase NMB0184 [imported] - Neisseria meningitidis  
 C:Species: Neisseria meningitidis  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: A81229  
 R:Ketterlin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Bisen, J.A.  
 Hickey, E.K.; Hatt, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 K. J.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Piazza, M.  
 Science 287, 1809-1815, 2000  
 A:Accession: A81229  
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A:Reference number: AB1000; MUID:20175755; PMID:10710307  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-394 <TEF>  
 A:Cross-references: GB:AB002375; GB:AB002098; NID:97225394; PIDN:AA640641.1; PID:9722540;  
 A:Experimental source: serogroup B, strain MC58  
 C:Genetics:  
 A:Gene: NMB0184  
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 52.4%; Score 1040.5; DB 2; Length 394;  
 Best local similarity 55.3%; Pred. No. 7, 6e-71;  
 Matches 215; Conservative 47; Mismatches 126; Indels 1; Gaps 1;  
 QY 4 ICIGATGSGITLVDVARRHPDKYQVVALTANGNIDALYEQCLAHPEYAVVMSKVA 63  
 DB 6 ILLANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 65  
 QY 64 EFKRIASPVADIKVLGSGEALQVATLENDVYMAIVGAAGLLPTLAAKAGKT 123  
 DB 64 EFKRIASPVADIKVLGSGEALQVATLENDVYMAIVGAAGLLPTLAAKAGKT 123  
 QY 124 ANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 183  
 DB 124 ANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 184  
 QY 125 ANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 184  
 DB 125 ANKEALVMSGQIFMOAVSDSGAVLIPIDSEHNAIFQCM--PAGYTPGHTAKQAR--- 184  
 QY 184 GPERTPETLSVTPDQAVAHFPMKDMGRKISVDSATMMNKGLELLACLLFENNEPOIE 243  
 DB 184 GPERTPETLSVTPDQAVAHFPMKDMGRKISVDSATMMNKGLELLACLLFENNEPOIE 244  
 QY 185 GPFLLADLAFTRITPAQAVKHPMWRMGKISVDSATMMNKGLELLACLLFENNEPOIE 244  
 DB 185 GPFLLADLAFTRITPAQAVKHPMWRMGKISVDSATMMNKGLELLACLLFENNEPOIE 244  
 QY 244 VVHPSGIIHSMVDYVDSVLAQKGNPDMPPIAHAMAMPFRDGVAPLDIFEGHMDP 303  
 DB 245 VVHPSGIIHSMVDYVDSVLAQKGNPDMPPIAHAMAMPFRDGVAPLDIFEGHMDP 304  
 QY 304 EKPDLPKPPCRLAYEALIKSGGIMPTVLNANEIAYEAFLEBEVKFTDIYVIRSMQAF 363  
 DB 305 EKPDLPKPPCRLAYEALIKSGGIMPTVLNANEIAYEAFLEBEVKFTDIYVIRSMQAF 364  
 QY 364 KPDAGSLBLVLAQDQAREVARDIITLVA 392  
 DB 365 FSDGIGDIGLADQADARTRAAFITGL 393

RESULT 7  
 AF052  
 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.-) [imported] - *Salmonella enterica*  
 A:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: AF0529  
 R:Parfitt, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
 Th. T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Cougle, S.; O'Garra, P.  
 Nature 413, 848-852, 2001  
 A:Accession: AF0529  
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608



A:Accession: AF0529  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-398 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CMD08678.1; PID:g16501501; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: dxt  
 C:Superfamily: conserved hypothetical protein HI0807  
 C:Keywords: oxidoreductase

Query Match 51.7%; Score 1027; DB 2; Length 398;  
 Best local similarity 55.8%; Pred. No. 8e-70;  
 Matches 223; Conservative 49; Mismatches 104; Indels 24; Gaps 7;

QY 1 MKGICILGATGSGTLDVVAHHPDKYQVVALTANGNIDALYECCLAHHPYAVVMES 60  
 DB 1 MKGICILGATGSGTLDVVAHHPDKYQVVALTANGNIDALYECCLAHHPYAVVMES 60  
 QY 61 KVAEFKQRIASPVDIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAAGAAGT 120  
 DB 61 KVAEFKQRIASPVDIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAAGAAGT 120  
 QY 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAAGLPTLAIRAGET 119  
 DB 61 SAQKIMQOHG-SKTEVLSSQACERAAALDEGVHVAIVGAAGLPTLAIRAGET 119  
 QY 121 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-AGYTP---GHT 170  
 DB 121 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-AGYTP---GHT 170  
 QY 120 ILLANKESILVTCGRLEPMDEYKSNARLLPYDSEHNAIFQSLPQIGHNIGYADLEQNGVT 179  
 DB 120 ILLANKESILVTCGRLEPMDEYKSNARLLPYDSEHNAIFQSLPQIGHNIGYADLEQNGVT 179  
 QY 171 AKGARILLTAGSGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 230  
 DB 171 AKGARILLTAGSGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 230  
 QY 180 S-----ILLTGSGGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 234  
 DB 180 S-----ILLTGSGGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 234  
 QY 231 ACCLFNMEDPQIEVVIHPSIIHSKVDYVDSVLAQMGPNDRKPIAHAMAFPERFDSGV 290  
 DB 231 ACCLFNMEDPQIEVVIHPSIIHSKVDYVDSVLAQMGPNDRKPIAHAMAFPERFDSGV 290  
 QY 235 ARLTFNANARQMEVLHPQSVHSHVRQDGSVLAQLGEPDRKPIAHAMAFPERFDSGV 294  
 DB 235 ARLTFNANARQMEVLHPQSVHSHVRQDGSVLAQLGEPDRKPIAHAMAFPERFDSGV 294  
 QY 291 APADIEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVKET 350  
 DB 291 APADIEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVKET 350  
 QY 295 QPDLFCILSLATYSADYQRYPCILKAMEAFEGQAATTAANAEITVAALFACQIRFT 354  
 DB 295 QPDLFCILSLATYSADYQRYPCILKAMEAFEGQAATTAANAEITVAALFACQIRFT 354  
 QY 351 DIA---VILERSMAQFKPDAGSLVLAQADQAREVAR 386  
 DB 351 DIA---VILERSMAQFKPDAGSLVLAQADQAREVAR 386  
 QY 355 DIAGMLAVLER-NDLHP---ASVDVLAQVDAIAREVAR 390  
 DB 355 DIAGMLAVLER-NDLHP---ASVDVLAQVDAIAREVAR 390

## RESULT 8

1-deoxy-D-xylose 5-phosphate reductoisomerase VC2254 [imported] - Vibrio cholerae [str

C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: D82099  
 C:Residues: 1-402 <HEI>

R:Reidberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gilm, M.L.; Dodson, R.J.;

Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Frazer, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AB2035; MUID:20406833; PMID:10952301

A:Accession: D82099

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-402 <HEI>

A:Cross-references: GB:AB004297; GB:AB003852; MID:g9656810; PIDN:AAFP95398.1; GSPDB:GN001

C:Genetics:

A:Experimental source: serogroup O1, strain N16961; biotype El Tor

A:Gene: VC2254

A:Map position: 1

C:Superfamily: conserved hypothetical proteain HI0807

Query Match 51.5%; Score 1023; DB 2; Length 402;  
 Best local similarity 52.7%; Pred. No. 1.6e-69;  
 Matches 214; Conservative 65; Mismatches 103; Indels 24; Gaps 4;

QY 1 MKGICILGATGSGTLDVVAHHPDKYQVVALTANGNIDALYECCLAHHPYAVVMES 60  
 DB 1 MKGICILGATGSGTLDVVAHHPDKYQVVALTANGNIDALYECCLAHHPYAVVMES 60

QY 61 KVAEFKQRIASPVDIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAAGAAGT 120  
 DB 61 KVAEFKQRIASPVDIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAAGAAGT 120  
 QY 121 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-PAGYTPGHTAKOAR----- 175  
 DB 121 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-PAGYTPGHTAKOAR----- 175  
 QY 120 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-PAGYTPGHTAKOAR----- 172  
 DB 120 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-PAGYTPGHTAKOAR----- 172  
 QY 176 -----RILLTAGSGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 228  
 DB 176 -----RILLTAGSGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 228  
 QY 173 LSGHIDHILLTGSGGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 232  
 DB 173 LSGHIDHILLTGSGGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 232  
 QY 229 IEACILFNMEDPQIEVVIHPSIIHSKVDYVDSVLAQMGPNDRKPIAHAMAFPERFDS 288  
 DB 229 IEACILFNMEDPQIEVVIHPSIIHSKVDYVDSVLAQMGPNDRKPIAHAMAFPERFDS 288  
 QY 233 IEAKLFTSREQLKVLHPQSVHSHVRQDGSVLAQLGEPDRKPIAHAMAFPERFDS 292  
 DB 233 IEAKLFTSREQLKVLHPQSVHSHVRQDGSVLAQLGEPDRKPIAHAMAFPERFDS 292  
 QY 289 GVAPLDIEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVK 348  
 DB 289 GVAPLDIEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVK 348  
 QY 293 GVPALDIEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVK 352  
 DB 293 GVPALDIEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVK 352  
 QY 349 FTDIAVILERSMAQFKPD---AGSLEVLVAQADQAREVAR 390  
 DB 349 FTDIAVILERSMAQFKPD---AGSLEVLVAQADQAREVAR 390  
 QY 353 FTDIALNDQVLSKVCATNTQLHCRDLESILDTMAHFAHQA 398  
 DB 353 FTDIALNDQVLSKVCATNTQLHCRDLESILDTMAHFAHQA 398

## RESULT 9

1-deoxy-D-xylose 5-phosphate reductoisomerase (EC 1.1.1.-) [imported] - Yersinia pestis

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AG0128

R:Phillips, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.

denk Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, Rutterford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AG0128

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-398 <KIR>

A:Cross-references: GB:AL590842; PIDN:CAG89890.1; PID:g15979115; GSPDB:GN00175

A:Gene: dxt

C:Superfamily: conserved hypothetical proteain HI0807

C:Keywords: oxidoreductase

Query Match 51.4%; Score 1020; DB 2; Length 398;  
 Best local similarity 55.0%; Pred. No. 2.7e-69;  
 Matches 219; Conservative 51; Mismatches 120; Indels 8; Gaps 3;

QY 1 MKGICILGATGSGTLDVVAHHPDKYQVVALTANGNIDALYECCLAHHPYAVVMES 60  
 DB 1 MKGICILGATGSGTLDVVAHHPDKYQVVALTANGNIDALYECCLAHHPYAVVMES 60  
 QY 61 KVAEFKQRIASPVDIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAAGAAGT 120  
 DB 61 KVAEFKQRIASPVDIKVLSGSEALQVATLENDVTMAAIVGAAGLPTLAAGAAGT 120  
 QY 61 S-AKSLRULLEGGSDTVVSGTACELALDDVDQVMAIVGAGLPTLAIRAGET 119  
 DB 61 S-AKSLRULLEGGSDTVVSGTACELALDDVDQVMAIVGAGLPTLAIRAGET 119  
 QY 121 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-AGYTPGHTAKO 174  
 DB 121 VLLANKALVMSGQIFMGAVSDSGAVLLPIDSEHNAIFQCM-AGYTPGHTAKO 174  
 QY 120 VLLANKESILVTCGRLEPMDEYKSNARLLPYDSEHNAIFQSLPQIGHNIGYADLEQNGVT 178  
 DB 120 VLLANKESILVTCGRLEPMDEYKSNARLLPYDSEHNAIFQSLPQIGHNIGYADLEQNGVT 178  
 QY 175 RILLTAGSGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 234  
 DB 175 RILLTAGSGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 234  
 QY 179 STIILTGSGGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 238  
 DB 179 STIILTGSGGPRRTPIETLSVTPDQAVAHKXMGKISVDSATMANKGLEIE 238  
 QY 235 FNMEDPQIEVVIHPSIIHSKVDYVDSVLAQMGPNDRKPIAHAMAFPERFDS 294  
 DB 235 FNMEDPQIEVVIHPSIIHSKVDYVDSVLAQMGPNDRKPIAHAMAFPERFDS 294  
 QY 239 FNASAQIEVVLHPQSVHSHVRQDGSVLAQLGEPDRKPIAHAMAFPERFDS 298  
 DB 239 FNASAQIEVVLHPQSVHSHVRQDGSVLAQLGEPDRKPIAHAMAFPERFDS 298  
 QY 295 IEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVK 354  
 DB 295 IEVGHMDPEKDLKRFPCILRAYEAIKSGIMPTVLAANEAIVAEFLNEBVK 354



A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDN:BA10183.1; PID:g100155  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Superfamily: conserved hypothetical protein HI0807

Query Match 45.0%; Score 892.5; DB 2; Length 394;  
 Best Local Similarity 47.3%; Pred. No. 1,1e-59;  
 Matches 183; Conservative 63; Mismatches 126; Indels 13; Gaps 3;

QY 1 MKGICILGATGSGVSTLDVVAHPDKQYVALTANGNIDALYEQCLAHPEYAVVMS 60  
 DB 2 VKRISILGSGTSGIGTDLIVTHHPAFQYVGLAAGNVALIAQVAEPEEIVAIROAE 61  
 QY 61 KVAEPRQIAASPVADIKVLGSEALQOVATLENVDYMAAIVGAAGLLPTLAAGA 120  
 DB 62 KVEDIAAAVAELTDYQPVYVBEQVEYARVGADESATVIGVCGLLPTLAAGA 121  
 QY 121 VLLANKREALVMSGOIFMCAVSDGAVLPIIDSEHNAIPQCM--PAGYTPGHTAKAARI 177  
 DB 122 IALANKETILAGAPVLPVEXKMGVYLLPADSEHSAIPQCLGVEG-----GLARI 173  
 QY 178 LLTASGAPRRRTPIETLSVTPDQAVAHKMKKISVDSATMKKGLIELIACILPMM 237  
 DB 174 ILLASGAFRRDIPVEELPVTVDALKHPMWSGKRTITDSATLANKGLEVEIAHYLFC 233  
 QY 236 EPDOIEVVIHPOSIIHSMVDYVDSGLAQMGNPMKPTIAHAMAPEREDSGVAPLIFE 297  
 DB 234 DTDHIDIVHPOSIIHSLIEVDSTVLAQCMFDRMLPLVLSMPERITDMEPLDYK 293  
 QY 298 VGHMDEKDLKRPFCRLIAYEAISGSGIMPTVLANAEIYEAFLNEVKFTDIAVIE 357  
 DB 294 AGSLIFREPDHDKYPCMQCLAYAGGAGAMPALNANQAAVALFOEKISFLDIPRLIE 353  
 QY 358 RSMAGKPKPDAGS--LEIVLQADQDAR 382  
 DB 354 KTCDDLVGQNTASPDLETLLADQMAR 380

## RESULT 13

F84957

1-deoxy-D-xylulose 5-phosphate reductoisomerase [imported] - Buchnera sp. (strain APS)

C:Species: Buchnera sp.  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: F84957  
 R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A  
 A:Reference number: AB4930; MUID:20445173; PMID:1093077

A:Accession: F84957

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-398 <STO>

A:Cross-references: GB:AP000398; GSPDB:GN00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: dxr; BU235

C:Superfamily: conserved hypothetical protein HI0807

Query Match 44.9%; Score 891; DB 2; Length 398;  
 Best Local Similarity 45.7%; Pred. No. 1.4e-59;  
 Matches 182; Conservative 83; Mismatches 127; Indels 6; Gaps 3;

QY 1 MKGICILGATGSGVSTLDVVAHPDKQYVALTANGNIDALYEQCLAHPEYAVVMS 60  
 DB 1 MKKITTLSGSGIGTDLIVTHHPAFQYVGLAAGNVALIAQVAEPEEIVAIROAE 61  
 QY 61 KVAEPRQIAASPVADIKVLGSEALQOVATLENVDYMAAIVGAAGLLPTLAAGA 120  
 DB 61 SAHILAKKLKHSKI-NTQVLTGKIDICALAALEETDHVISAIVGAGLLPTLSAHAGKT 119  
 QY 121 VLLANKREALVMSGOIFMCAVSDGAVLPIIDSEHNAIPQCMAGYTP--GHTA---KQAR 175  
 DB 120 ILLANKREALVMSGOIFMCAVSDGAVLPIIDSEHNAIPQCMAGYTP--GHTA---KQAR 175

QY 176 RILITAGSGPPRRPIETLSVTPDQAVAHKMKKISVDSATMKKGLIELIACILF 235  
 DB 180 HVLITGGGPPYKSSDLSVTPDQACSHPMWLMGKISVDSATMKKGLIEVAEAKLF 239  
 QY 236 NMEPDQIEVVIHPOSIIHSMVDYVDSGLAQMGNPMKPTIAHAMAPEREDSGVAPL 295  
 DB 240 NALSEIKILIHPSIIHSMVQYVDSGLAQLSPDILRTISYVMSMPDRICTEVDYLN 299  
 QY 296 FEVGHMDEKDLKRPFCRLIAYEAISGSGIMPTVLANAEIYEAFLNEVKFTDIAV 355  
 DB 300 YKINMLTFEEDPFPQCLKXIAIDAFSGQASMTVLANAEIYEAFLNEVKFTDIAV 359  
 QY 356 IERSMAGKPKPDAGSLEIVLQADQDAREVARDIKTYV 393  
 DB 360 NMEILMSSCSPEKCIQDILHIDKVALIAKAKVSSLI 397

## RESI T 14

1-deoxy-D-xylulose 5-phosphate reductoisomerase [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Name: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C:Accession: AB2349

R:Kato, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, S.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

Nature 407, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-399 <KUS>

A:Cross-references: GB:BA000019; PIDN:BA876050.1; PID:g17133487; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: dxr; alr4351

C:Superfamily: conserved hypothetical protein HI0807

Query Match 43.6%; Score 865; DB 2; Length 399;  
 Best Local Similarity 44.8%; Pred. No. 1.3e-57;  
 Matches 180; Conservative 76; Mismatches 126; Indels 20; Gaps 5;

QY 1 MKGICILGATGSGVSTLDVVAHPDKQYVALTANGNIDALYEQCLAHPEYAVVMS 60  
 DB 2 VKRISILGSGTSGIGTDLIVSOYPPDQFRIVGLAAGSVEMLAQIRQFRFOIAISAE 61  
 QY 61 KVAEPRQIA--ASPVADIKVLGSEALQOVATLENVDYMAAIVGAAGLLPTLAAGA 117  
 DB 62 KLPALQAAIKLDDQPI---ILGGEAGVIEVARYGAEITVIGCGAGLLPTLAIEA 117  
 QY 118 KTVLLANKREALVMSGOIFMCAVSDGAVLPIIDSEHNAIPQCM--PAGYTPGHTAKOA 174  
 DB 118 GKDILANKETILAGAPVLPVEXKMGVYLLPADSEHSAIPQCLGVEG-----GL 169  
 QY 175 RRIILITAGSGPPRRPIETLSVTPDQAVAHKMKKISVDSATMKKGLIELIACIL 234  
 DB 170 KKLITLTSAGGAFRRDIPVEELPVTVDALKHPMWSGKRTITDSATLANKGLEVEIAHYL 229  
 QY 235 NMEPDQIEVVIHPOSIIHSMVDYVDSGLAQMGNPMKPTIAHAMAPEREDSGVAPL 294  
 DB 230 FGLDQDIEIYHPOSIIHSLIEVDSTVLAQCMFDRMLPLVLSMPERITDMEPLDYK 289  
 QY 295 IFVGHMDEKDLKRPFCRLIAYEAISGSGIMPTVLANAEIYEAFLNEVKFTDIAV 354  
 DB 290 LVKAGNLTFRPDDQKPCQCLAYAGGAGAMPALNANQAAVALFOEKISFLDIPRLIE 349  
 QY 355 IERSMAGKPKPDAG--SLEIVLQADQDAREVARDIKTYV 394  
 DB 350 CIELVCDRHQNDNCANPDLDDIILAADQASQEVLTATKTLAS 391

RESI T 15

HB7486

1-deoxy-D-xylulose 5-phosphate reductoisomerase [imported] - *Caulobacter crescentus*  
C. Species: *Caulobacter crescentus*

C:Species: *caulobacter crescentus*  
C:Date: 20-Apr-2001 #ext chance 10-May-2001

C:\Accession: H07485  
#sequence\_revision 20-Apr-2001 #exec\_change 10-May-2001

R; Nierman, W. C.; Fe

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.E.; Kolor

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

3. Reference number: MGITD.31173699: BMTD.11359647

Q: Accession: H874B6  
N, Reference Number: 80/275, MUD: 211/3030, FID: 11225071

A: Status: preliminary

A;Molecule type: DNA

A;Residues: 1-399 &lt;ST

A: Cross-references: GB:

C:\Genetics:

C:SuperfamilyV: conserved hypothetical protein HT0807

[illegible]

Query Match 43.2%; Score 857.5; DB 2

Best Local Similarity 44.6%; Pred. No. 4.9e-57;

Matches 175; Conservative 77; Mismatches 125; Indels 15; Gaps 5;

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Db 10 VVLTGSGTGLSTLSLFEESGAPQVLTALTAGNVERLTBOARKWPSLAVIEDESRD 69

[illegible]

Db 70 DLRGLAGT---GVEAAAGADAVRDAAAM--GADWWSAIVGAAGLAPVAAAPRTGAVATL 125

00 134 ANKEBIIYMGOTEMOAVUSCSAVI:BITSEUNAIFOCMBVCYEDCHTAROBPTI:ETBSC 183

Db 126 ANKESLVCAGPALATIAKAGGCSVIVPDSEHSALFQYLQS-----ECAHRVSLILITNSG 180

184 GREFETDTFTTSSYMPDQAVAHPTWMPRTSYMSATMNKGIETFACTJENMPDQTE 243

Db 181 GPFRWDKAMARATPEQAIHPNWSGAKISVDSATMTNKGLEMIKSYLFATPEEDVD 240

244 VTIHPDSTIHSMDVYDNGSVTAOMGNPDMETPTAHAMAWPERFDSQVAPIDTFEVCMDP 303

Db 241 VVHHQSVIHSINVEYDGSTLAQIGPDMRAPACAPWPDRLPWAPRLDLAAVGLTF 3000

304 EKPDLEKPECLRLAYEAIKSGGIMPTVLNAANEIAYEAFINEBKETDIAYII-----ER- 358

Db 301 ESPVTERFPAIGIAREBALRLOGGAPAPMANNAEVAVAALFIDRIGELDIAGAVAGTLERM 360

OV 359 -SMAOFKPPDAGSLELVLQADODAREVARDIT 389

361 NSLGDISVAESDAVETAMLIDGSARRIAEVV 3922

Search completed: January 29, 2004, 15:55:30  
Job time : 12.5928 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 7.41938 Seconds

(without alignments)  
2497.314 Million cell updates/sec

Title: US-09-941-947a-8

Perfect score: 1985

Sequence: 1 MGICICIGATGSGISGVTLDV.....LQADQDAEVADIIKTLVA 394

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Score  | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1          | 1215.5 | 61.2        | 396    | 1     | DXR_PSEAB   |
| 2          | 1091.5 | 55.0        | 393    | 1     | DXR_RALSO   |
| 3          | 1062.5 | 53.5        | 402    | 1     | DXR_VIBVU   |
| 4          | 1052.5 | 53.0        | 394    | 1     | DXR_NEIMA   |
| 5          | 1047.5 | 52.7        | 401    | 1     | DXR_VIRPA   |
| 6          | 1045.5 | 52.6        | 398    | 1     | DXR_ECOLI   |
| 7          | 1044.5 | 52.6        | 398    | 1     | DXR_ECO57   |
| 8          | 1042.5 | 52.5        | 405    | 1     | DXR_PACNU   |
| 9          | 1040.5 | 52.4        | 394    | 1     | DXR_NEIMA   |
| 10         | 1034.5 | 52.1        | 398    | 1     | DXR_SALTY   |
| 11         | 1027.5 | 51.7        | 398    | 1     | DXR_SALTY   |
| 12         | 1023.5 | 51.5        | 402    | 1     | DXR_VIRCH   |
| 13         | 1020.5 | 51.4        | 398    | 1     | DXR_YERPE   |
| 14         | 1007.5 | 50.7        | 397    | 1     | DXR_HARIE   |
| 15         | 999.5  | 50.4        | 398    | 1     | DXR_ZYMON   |
| 16         | 977.5  | 49.2        | 396    | 1     | DXR_XANAC   |
| 17         | 973.5  | 49.0        | 396    | 1     | DXR_XANCP   |
| 18         | 941.5  | 47.4        | 396    | 1     | DXR_XYLP    |
| 19         | 892.5  | 45.0        | 394    | 1     | DXR_SYNY3   |
| 20         | 892.5  | 44.9        | 398    | 1     | DXR_BUCAP   |
| 21         | 891.5  | 44.9        | 398    | 1     | DXR_BUCAP   |
| 22         | 887.5  | 44.7        | 394    | 1     | DXR_BUCAP   |
| 23         | 886.5  | 44.6        | 402    | 1     | DXR_CLOPE   |
| 24         | 868.5  | 43.7        | 382    | 1     | DXR_BACHD   |
| 25         | 865.5  | 43.6        | 399    | 1     | DXR_ANASP   |
| 26         | 857.5  | 43.2        | 399    | 1     | DXR_CAUCR   |
| 27         | 854.5  | 43.0        | 385    | 1     | DXR_CLOAB   |
| 28         | 854.5  | 43.0        | 391    | 1     | DXR_BHIME   |
| 29         | 849.5  | 42.8        | 397    | 1     | DXR_AGRIS   |
| 30         | 837.5  | 42.2        | 382    | 1     | DXR_THERT   |
| 31         | 818.5  | 41.2        | 397    | 1     | DXR_WIGBR   |
| 32         | 818.5  | 41.2        | 477    | 1     | DXR_ARATH   |
| 33         | 816.5  | 41.1        | 380    | 1     | DXR_LISMO   |

| Result ID | Score | Query Match | Length | DB ID | Description |
|-----------|-------|-------------|--------|-------|-------------|
| 1         | 812.5 | 40.9        | 380    | 1     | DXR_LISIN   |
| 2         | 798.5 | 40.2        | 388    | 1     | DXR_BACSU   |
| 3         | 743.5 | 37.4        | 390    | 1     | DXR_FUSNN   |
| 4         | 741.5 | 37.3        | 475    | 1     | DXR_MENPI   |
| 5         | 707.5 | 35.6        | 379    | 1     | DXR_CHLPP   |
| 6         | 698.5 | 35.2        | 396    | 1     | DXR_DEIRA   |
| 7         | 696.5 | 35.1        | 418    | 1     | DXR_SMRCC   |
| 8         | 692.5 | 34.9        | 392    | 1     | DXR_CORGL   |
| 9         | 691.5 | 34.8        | 413    | 1     | DXR_MYCTU   |
| 10        | 684.5 | 34.5        | 379    | 1     | DXR_CHLMT   |
| 11        | 681.5 | 34.3        | 406    | 1     | DXR_MYCLE   |
| 12        | 646.5 | 32.6        | 379    | 1     | DXR_CHLTR   |

## ALIGNMENTS

| Result ID | Score | Query Match | Length | DB ID | Description |
|-----------|-------|-------------|--------|-------|-------------|
| 1         | 812.5 | 40.9        | 380    | 1     | DXR_LISIN   |
| 2         | 798.5 | 40.2        | 388    | 1     | DXR_BACSU   |
| 3         | 743.5 | 37.4        | 390    | 1     | DXR_FUSNN   |
| 4         | 741.5 | 37.3        | 475    | 1     | DXR_MENPI   |
| 5         | 707.5 | 35.6        | 379    | 1     | DXR_CHLPP   |
| 6         | 698.5 | 35.2        | 396    | 1     | DXR_DEIRA   |
| 7         | 696.5 | 35.1        | 418    | 1     | DXR_SMRCC   |
| 8         | 692.5 | 34.9        | 392    | 1     | DXR_CORGL   |
| 9         | 691.5 | 34.8        | 413    | 1     | DXR_MYCTU   |
| 10        | 684.5 | 34.5        | 379    | 1     | DXR_CHLMT   |
| 11        | 681.5 | 34.3        | 406    | 1     | DXR_MYCLE   |
| 12        | 646.5 | 32.6        | 379    | 1     | DXR_CHLTR   |

DR EMBL: A604785; AAG07038.1; --  
 DR PIR: E83188; E83188.  
 DR HAMAP: MF\_00183; -- 1.  
 DR InterPro: IPR003821; DXP\_reductoisom.  
 DR Pfam: PF02670; DXP\_reductoisom; 1.  
 DR TIGRfams: TIGR00243; Dxr; 1.  
 DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
 FT NP BIND 10 NADPH (POTENTIAL).  
 FT NP BIND 17  
 SQ SEQUENCE 396 AA; 42508 MW; 2ECB9D8A94A9CAD7 CRC64;

Query Match 61.2%; Score 1215.5; DB 1; Length 396;  
 Best Local Similarity 64.1%; Pred. No. 9,3e-80;

Matches 245; Conservative 48; Mismatches 88; Indels 1; Gaps 1;

QY 4 IITGATSGIGSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESYA 63  
 DB 7 ISVLATGSIIGSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVESYA 66  
 QY 64 EKKRIASPVADIKVLSGSEALQOVATLENDVTMAAIVGAGILPTIAAKAKGTVL 123  
 DB 67 ALQGSIAAAGIR-TEVLFGEQALCEVASAPEVDYMAAIVGAGILPTIAAVERGKVTLL 125  
 QY 124 ANKALVMSGQIFMQAVSDGAVLLPDSSENAIFQCPAGTPTHTAKQARRILLTASG 183  
 DB 126 ANKALVMSGALFMQAVRSQAVLLPDSSENAIFQCPAGTPTHTAKQARRILLTASG 185  
 QY 184 GFFRRTPITLSSVTPDQAVAHHPKMDGKRTISVDSATMANKGLEIEACLLFNMEDQIE 243  
 DB 186 GFFRRTPITLSSVTPDQAVAHHPKMDGKRTISVDSATMANKGLEIEACLLFNMEDQIE 245  
 QY 244 VVIHQSIHSMWVYDGSVLAQMGKPNMRTPIAHMAMPFERPDSGVAFLDIFEGANDF 303  
 DB 246 VVIHQSIHSMWVYDGSVLAQMGKPNMRTPIAHMAMPFERPDSGVAFLDIFEGANDF 305  
 QY 304 EKKDLPKPCPLATYEALKSGIMFTVNAANIEVEAFLEVEFTTIAVIERSMQF 363  
 DB 306 QRPDQRPFCPLATYEALKSGIMFTVNAANIEVEAFLEVEFTTIAVIERSMQF 365  
 QY 364 KPDAGSLLEVLQADQAREVA 385  
 DB 366 AVTAVESLDQVLAADRRARSA 387

RESULT 2  
 DNR\_RALSO STANDARD; PRT; 393 AA.

AC Q8XZ15;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).  
 GN DXR OR RSC1410 OR RS05282.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11030;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salenouat M., Gerin S., Artiguenave F., Gouzy J., Mangenot S., Ariat M., Billault A., Brothier P., Camus J.C., Catolico L., Chandler M., Choise N., Claudel-Renard C., Cunne S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Sigler P., Trebault P., Whalen K., Wincker P., Levy M., Weissbach J., Boucher C.A.;  
 RA "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RT Nature 415:497-502(2002).  
 CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (by similarity).  
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) +

CC = 1-deoxy-D-xyulose 5-phosphate + NADPH.  
 CC -1- PATHWAY: Nomevalonate terpenoid biosynthesis pathway; second step

CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL: A646064; CAD15112.1; --

DR HAMAP: MF\_00183; -- 1.  
 DR InterPro: IPR003821; DXP\_reductoisom.

DR Pfam: PF02670; DXP\_reductoisom; 1.  
 DR TIGRfams: TIGR00243; Dxr; 1.

DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
 FT NP BIND 7 NADPH (POTENTIAL).  
 FT NP BIND 14  
 SQ SEQUENCE 393 AA; 41763 MW; 5A955A7C8F32E9D CRC64;

QY Match 55.0%; Score 1091.5; DB 1; Length 393;  
 Best Local Similarity 58.2%; Pred. No. 6.7e-71;

M Chs 227; Conservative 49; Mismatches 109; Indels 5; Gaps 3;

QY 1 MKGICITGATSGIGSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVES 60  
 DB 1 MMRITVLGATSGIGSTLDVVAHHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWVES 60  
 QY 61 KVAEKKRIASPVADIKVLSGSEALQOVATLENDVTMAAIVGAGILPTIAAKAKGT 120  
 DB 61 KVAEKKRIASPVADIKVLSGSEALQOVATLENDVTMAAIVGAGILPTIAAKAKGT 119  
 QY 121 VILANKALVMSGQIFMQAVSDGAVLLPDSSENAIFQCPAGTPTHTAKQARRILLTASG 179  
 DB 120 VILANKALVMSGALFMQAVRSQAVLLPDSSENAIFQCPAGTPTHTAKQARRILLTASG 176  
 QY 180 TASGPFRRTPITLSSVTPDQAVAHHPKMDGKRTISVDSATMANKGLEIEACLLFNMED 239  
 DB 177 TASGPFRRTPITLSSVTPDQAVAHHPKMDGKRTISVDSATMANKGLEIEACLLFNMED 236  
 QY 240 DQLEVVHPOSIIHSMWVYDGSVLAQMGKPNMRTPIAHMAMPFERPDSGVAFLDIFEGANDF 299  
 DB 237 DQLEVVHPOSIIHSMWVYDGSVLAQMGKPNMRTPIAHMAMPFERPDSGVAFLDIFEGANDF 296  
 QY 300 HMDPEKDLKRPCLATYEALKSGIMFTVNAANIEVEAFLEVEFTTIAVIERSMQF 359  
 DB 297 HMDPEKDLKRPCLATYEALKSGIMFTVNAANIEVEAFLEVEFTTIAVIERSMQF 356  
 QY 360 MADQKPDAGSLLEVLQADQAREVA 389  
 DB 357 LARTALGPADSLDTPFAADQARRARHYI 386

REST T 3  
 DNR\_TBYU STANDARD; PRT; 402 AA.

AC Q8DBF5;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).  
 GN DXR OR VY11866.  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim U.J., Moon Y.H., Jeong H.,

CC acteria;Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
ON TBI\_TaxID=65699;  
RN []  
RP JGNCES FROM N.A.  
RC TRAIN=22491 / Serogroup A / Serotype 4A;  
RX SBLINE=2022255; PubMed=10761919;  
RA Arkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
R Lee S.H.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
R Wiles R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,  
R Ngels K., Leather S., Moule S., Mungall K., Quail M.A.,  
R Ngardean M.A., Rutherford K.W., Simmonds W., Skelton J.,  
R Hitched S., Spratt B.G., Bartell B.G.;  
RT Complete DNA sequence of a serogroup A strain of Neisseria  
RT meningitidis 22491."  
RL ture 404:502-506(12000).  
CC - FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction  
CC of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythrulose  
CC 4-phosphate (MEP) (By similarity).  
CC - CATALYTIC ACTIVITY: 2-C-methyl-D-erythrulose 4-phosphate + NADP(+) =  
CC - 1-deoxy-D-xylulose 5-phosphate + NADPH.  
CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second  
CC step.  
CC - SIMILARITY: BELONGS TO THE DXR FAMILY.  
-----  
CC his SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC te European Bioinformatics Institute. There are no restrictions on its  
CC ie by non-profit institutions as long as its content is in no way  
CC diffied and this statement is not removed. Usage by and for commercial  
CC ities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC BL; ALI62752; CAB83399.1; .  
DR FR; B82000; B82000.  
DR MAP; MF 00183; -! 1.  
DR lePro; IPR003821; DXP\_redoxomase.  
DR nam; PR02670; DXP\_reductoisom; 1.  
DR GRFams; TIGR00243; Dxrt; 1.  
DR loprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
DR BIND 9 NADPH (POTENTIAL).  
FR QUNCE 394 AA; 41912 MW; BSD361B224806DB CRC64;  
One March 53.0%; Score 1052.5; DB 1; Length 394;  
Bes Local Similarity 55.5%; Pred. No. 41e-68;  
Mat es 216; Conservative 49; Mismatches 123; Indels 1; Gaps 1;  
  
QY 4 ICIVGATGSIGVSTILDVAARRHPDKYQVAVLTANGNTDALLYBOCLAHHEPERAYVMESKYA 63  
Db 6 LFIIGSGISGESTLDVVSRRPEKRFVPALHGKQVEKLAACCTFFREYAVVADAHHAA 65  
QY 64 EFQRRIASVYADIKXUSGESEALQOVTLNVDPTVMAIYGAGLLFTLAAAKAGKTLYL 123  
Db 66 RLEMLLRDGT-TQVHGAQLPDVVASADEVSGMCALYGAVALPALAMAOMCKITLYL 124  
QY 124 ANKEALWMSQIFMQAVSDGAVLLPIDSEHNAIFQCMAPGYTPGHAKQARRILLTASG 183  
Db 125 ANKETLVVSGALLFMETARANGAAVLFPDSEHNAVGVLPEDYTRLNEHGIASIILTASG 184  
QY 184 GPFRRTIELTSLSTTPQOAVANPEMDGRKITSYDSATPMARKGLERACILEFNNEPPDIE 243  
Db 185 GFPITLADLNFDSTIPDQAVKHPEWRGKRISYDSATMANKGLEIRAHWLNFPCPKDIE 244  
QY 244 VVIHQSIIHSMVDYVVGSVLAQMGNNDMTPIAHAMWERPDSGVAPLDIFEVGMDF 303  
Db 245 VVIHQSVIHSMYRYRGSVLAQMGNDMTPIAYCGLPERIDSGVGDIDFDALSALT 304  
QY 304 EKPDLPFCPLRLAYELKGGIMPTYLANAANEIYAZVAFNEBEKTFDIYIIERSMAOF 363  
Db 305 QKPEDFDFPFLKAIYEMNNGAAPCYLANAEAAVAALFDGQIKFTFDIAKTVAHCLSOD 364  
QY 364 KPDDAGSELVLAGDDCARVARDIITKL 392



DB 365 FSDGIGDGLLADQARTRARARITGLT 393

RESULT 5  
DXR\_VIBPA STANDARD: PRT: 401 AA.

AC 087ME3;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE 1-deoxy-D-xylose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxylose-5-phosphate reductoisomerase).  
GN DXR OR VP2312  
OS *Vibrio parahaemolyticus*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrrio.  
OX NCBI\_TaxID=670;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RIMD 2210633 / Serotype O3.K6;  
RX MEDLINE=22508454; PubMed=12620739;  
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K., Iijima Y., Najiima M., Nakano M., Yamashita A., Kubota Y., Kimura S., Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
RA "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*.";  
RT Lancet 361:743-749(2003)  
RL  
CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xylose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (by similarity).  
CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xylose 5-phosphate + NADPH.  
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second step.  
CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AP005081; BAC60575.1; -.  
CC HAMAB: MF 00183; -; 1.  
KW Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
FT NP BIND 14 NADPH (POTENTIAL).  
SQ SEQUENCE 401 AA; 43109 MW; 470537A14305BD0 CRC64;

Query Match 52.7%; Score 1047; DB 1; Length 401;  
Best local Similarity 54.3%; Pred. No. 1e-67;  
Matches 222; Conservative 57; Mismatches 100; Indels 30; Gaps 5;

QY 1 MRCICIGTNGSGVTLQVNAHHPKXQVVALTAQGNIDALYEQGLAHREYAVVNES 60  
DB 1 MRCICIGTNGSGVTLQVNAHHPKXQVVALTAQGNIDALYEQGLAHREYAVVNES 60  
QY 61 KVAEFKQRI-ASPVADIKVLSSSEALQOVATLENDYVMAIYGAAGLLPTLAACK 119  
DB 61 KVAEFKQRI-ASPVADIKVLSSSEALQOVATLENDYVMAIYGAAGLLPTLAACK 119  
QY 61 AAVALQSEIHTTSP-NTETVLGGVDALCVVASLEVEDSVAAIYGAAGLLPTMAAVK 118  
DB 61 AAVALQSEIHTTSP-NTETVLGGVDALCVVASLEVEDSVAAIYGAAGLLPTMAAVK 118  
QY 120 TULANKKALVMSGOIFMAVSDSGAVLLPISSEHNAIQCPAGTFRHTAKQAR---- 175  
DB 120 TULANKKALVMSGOIFMAVSDSGAVLLPISSEHNAIQCPAGTFRHTAKQAR---- 175  
QY 119 RYLLANKKALVMSGOIFDAVEQYGAELLPVNSEHNAIFQCLP-----QOVQTNIG 169  
DB 119 RYLLANKKALVMSGOIFDAVEQYGAELLPVNSEHNAIFQCLP-----QOVQTNIG 169  
QY 176 -----RIILTASGPFRRPTIETLSVTPDQAVAPKMDMKRISVDSATMMNG 225  
DB 176 -----RIILTASGPFRRPTIETLSVTPDQAVAPKMDMKRISVDSATMMNG 225  
QY 170 RCMLDHGGISILLTSGGPFRRYADADSVTPAQAIAHPNMSGPKISVDSATMMNG 229  
DB 170 RCMLDHGGISILLTSGGPFRRYADADSVTPAQAIAHPNMSGPKISVDSATMMNG 229  
QY 226 LEIIEKCLFNMPEQIEVVIHQSIHKNVYVDSVLAQKGNPMPKPIIAHNAAMPR 285  
DB 226 LEIIEKCLFNMPEQIEVVIHQSIHKNVYVDSVLAQKGNPMPKPIIAHNAAMPR 285  
QY 230 LEYIEAKWLFNARDQIKVLIHQSVIHSWVYRDSVLAQKGPMPKPIIALTWSYPR 289  
DB 230 LEYIEAKWLFNARDQIKVLIHQSVIHSWVYRDSVLAQKGPMPKPIIALTWSYPR 289

QY 286 FDSGVADLDFEYGHMDFEKDILKRPCLRLAYBAIKSGGIMPTVLAANEIAYEALINE 345  
DB 290 VDAVGVKLEDFYQVGBLFTLQDPFARYPCCLALADACBQGHATTAANAEEVAVDAPLNN 349  
QY 346 EVKFTDIA----VIERSMAPQFPPDGSLELYLQADQAREYVARDIIX 390  
DB 350 RLGFDTIARINELVLIKRTASCTPEANNSLESLDNRMTALAEIR 398

RESI T 6  
DXR COLI STANDARD: PRT: 398 AA.

ID 24568; P77209;  
AC 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE 1-deoxy-D-xylose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxylose-5-phosphate reductoisomerase).  
GN DXR OR B0173.  
OS *Escherichia coli*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC STRAIN=KL2 / W3110;  
RX MEDLINE=98374274; PubMed=9707569;  
RA Takahashi S., Kuzuyama T., Matsumbe H., Seto H.;  
RT "A 1-deoxy-D-xylose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:9879-9884(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KL2 / MG1655;  
RX MEDLINE=9742617; PubMed=9278503;  
RA Blatter F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D., Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KL2 / W3110;  
RX Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T., Yamamoto Y., Inokuchi H., Inokuchi H., Hatada B., Fukuda R., Ichihara S., Mizuno T., Makino K., Nakata A., Yura T., Sangei G., Mitsuuchi K.;  
RT "Systematic sequencing of the *Escherichia coli* genome: analysis of the 4.0 - 6.0 min (189,987 - 281,416bp) region.";  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S. Duncan M., Allen B., Araujo R., Aparicio A., Chung E., Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurtz O., LaSakari D., Lew H., Lin D., Namath A., Oelner P., Roberts D., Davis R.W.;  
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-49 FROM N.A.  
RC MEDLINE=93077430; PubMed=1447125;  
RX Yamanaka K., Ogura T., Niki H., Hiraga S.;  
RT "Identification and characterization of the *smuA* gene, a suppressor of the *mukB* null mutant of *Escherichia coli*.";  
RL J. Bacteriol. 174:7517-7526(1992).  
RN [6]  
RP IDENTIFICATION.  
RX MEDLINE=96032851; PubMed=7567469;  
RA Borodovsky M., Weinlich J., Koonin E.V., Rudd K.E., Medigue C., Danchin A.;  
RT "Detection of new genes in a bacterial genome using Markov models for three gene classes.";

RL Nucleic Acids Res. 23:3554-3562(1995).  
 RN [7]  
 RP CHARACTERIZATION  
 FX MEDLINE=20123893; PubMed=10631325;  
 RA Radeyewicz T., Rohdich F., Mungstintawekul J., Hertz S., Kis K.,  
 RA Eisenreich W., Bacher A., Zenk M.H., Arigoni D.;  
 RT "Biosynthesis of terpenoids: 1-deoxy-D-xylulose-5-phosphate  
 RT reductoisomerase from *Escherichia coli* is a class B dehydrogenase.";   
 RL FEBS Lett. 465:157-160(2000).  
 RN [8]  
 RP CHARACTERIZATION AND MUTAGENESIS.  
 RX MEDLINE=20347905; PubMed=10787409;  
 RA Kuriyama T., Takahashi S., Takagi M., Seto H.;  
 RT "Characterization of 1-deoxy-D-xylulose 5-phosphate reductoisomerase,  
 RT an enzyme involved in isopentenyl diphosphate biosynthesis, and  
 RT identification of its catalytic amino acid residues";  
 RL J. Biol. Chem. 275:19928-19932(2000).  
 CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction  
 CC of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol  
 CC 4-phosphate (MEP).  
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)  
 CC = 1-deoxy-D-xylulose 5-phosphate + NADPH.  
 CC -1- COFACTOR: REQUIRES A DIVALENT CATION; MAGNESIUM; MANGANESE OR  
 CC COBALT.  
 CC -1- ENZYME REGULATION: INHIBITED BY POSITIDOMYCIN.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second  
 CC step.  
 CC -1- SUBUNIT: Homotetramer.  
 CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/sib.ch>).  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AB013300; BAA32426.1; -;  
 DR EMBL: AE000126; AAC73284.1; -;  
 DR EMBL: D83536; BAA7848.1; -;  
 DR EMBL: U70214; AAB08602.1; -;  
 DR EMBL: D13334; -; NOT ANNOTATED\_CDS.  
 DR PIR: B64741; E64741.  
 DR PDB: 1K5H; 27-FEB-02.  
 DR EcGene: BG12715; dxr.  
 DR HAMAP: MF\_00183; -; 1.  
 DR InterPro: IPR003821; DXR\_reductoisomase.  
 DR Pfam: PF02670; DXR\_reductoisom\_1.  
 DR TrEMBL: TIGR00243; Dxr; 1.  
 KW isoprenes biosynthesis; Oxidoreductase; NADP; Magnesium; Manganese;  
 KW Cofact; Complete proteome; 3D-structure.  
 KW NP BIND 7 14 NADPH (POTENTIAL).  
 FT MUTAGEN 153 14 G->D: DECREASE IN ACTIVITY.  
 FT MUTAGEN 153 153 H->Q: INCREASE IN KM.  
 FT MUTAGEN 209 209 H->Q: INCREASE IN KM.  
 FT MUTAGEN 231 231 E->K: DECREASE IN ACTIVITY.  
 FT MUTAGEN 257 257 H->Q: INCREASE IN KM.  
 FT FT 277 284 RPIIHTM -> VROLPPPW (IN REF. 3).  
 SQ SEQUENCE 398 AA; 43388 MW; 9B5326834AFF1207 CRC64;  
 Query Match 52.6%; Score 1045; DB 1; Length 398;  
 Best Local Similarity 54.9%; Pred No. 1.4e-67;  
 Matches 219; Conservative 56; Mismatches 118; Indels 6; Gaps 3;  
 QY 1 MKGICITAGTSGIVSTLDVVAHRBDKTVVALLTANGNIDALYECCIAHREYAVVMES 60  
 DB 1 MKQLTILSTGSGISTLDVVAHNEHFRRVVALVAGKVTYRVEECLEFSRYAAMDSEA 60  
 QY 61 KVAEKKQIIAASPVAIDIKVLSGSESLAQVATLENDVTMAALVAGAGLLPTLAACKAKT 120  
 DB 61 S-AKLLKTMLOOQGSRTFVLSQQAACDPAALIEDVDQVMAALVGAAGLLPTLAIRAGKT 119

```

QY 121 VILANKALVMSGGIIFQQAQSDSGAVLLPIDSINNAIFQCM--FAGYTPGHATAQAR---175
Db 120 ILANKASVTCGRLFPMDFADKQKSGAQQLLPVDSINNAIFQSI.PQSI.PQSI.GHNLGADLEQNGV 179
QY 176 RIILITAGGPPRRPTPIETSSVTPDQVAHPKMDMGKLSYDSTAATNNKGIELIEACILP 235
Db 180 SIILTSGCGPFRPTPIELDLATMTPTDQCRPHNNMSMKRI SYDSTAATNNKGILEYEAAML 239
QY 236 NMEPDQIEVYIHPSIIHSHWVDYDGSVLQWGMGNPDKRTPIAHAMWPERPDSGVAPADI 295
Db 240 NAGASQWQEVLIHPOSIVHSWRYDQSVLQLGEPDKRTPIAHMAMPNNKNSGVKELDP 299
Db 300 CKLSALTFEAPDYDRYFCLKILAMEAFRQQAATTAANAHEITVAALAAQOIRPTDIAL 359
QY 356 IESMAQFRPDDAGSLSLVLAQADDAEVAARDIIKTLVA 394
Db 360 NLSYLERMDREPCVDVVLSDANAAEVAKEVRLAS 398

RESULT:
7
DXR_E 357
ID CR ECO57 STANDARD; PRI: 398 AA.
AC 3ABIT;
AC 3-FEB-2003 (Rel. 41, Created)
DT 3-FEB-2003 (Rel. 41, Last sequence update)
DT 3-FEB-2003 (Rel. 41, Last annotation update)
DE -deoxy-D-xyululose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP
DE -decoisomerase) (1-deoxyxyululose-5-phosphate reductoisomerase).
GN KR Z0184 OR EC50175.
OS Sclerichia coli O157:H7.
OC Actetia; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Aerobacteriaceae; Escherichia.
OX ZBI_TaxID=83334;
1]
SEQUENCE FROM N.A.
RP TRAIN=O157:H7 / EDL933 / ATCC 700927;
RC EDLINE=21074935; PubMed=1120655;
RC zma N.T., Plunkett G., III, Burland V., Mau B., Glaesner J.D.,
RX jse D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA sfef G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA colbeck E.J., Davis N.W., Lim A., Dimalaria E.T., Potamoustis K.,
RA podaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Jlich R.A., Blattner P.R.;
RT genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL ture 409:529-533(2001).
1]
SEQUENCE FROM N.A.
RP TRAIN=O157:H7 / RIMD 0509952;
RC EDLINE=21156231; PubMed=11258796;
RX yashit T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA in C.-G., Ohtsubo H., Nakayama K., Murata T., Tanaka M., Itoe T.,
RA ida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,
RA hara S., Shiba T., Hattori M., Shinagawa H.;
RT complete genome sequence of enterohaemorrhagic Escherichia coli
RT 57:H7 and genomic comparison with a laboratory strain K-12."
RL JA Res. 8:11-22(2001).
CC FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction
CC of 1-deoxy-D-xyululose-5-phosphate (DXP) to 2-C-methyl-D-erythritol
CC 4-phosphate (MEP) (By similarity).
CC 1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)
CC = 1-deoxy-D-xyululose 5-phosphate + NADPH.
CC - PATHWAY: Nornomevalonate terpenoid biosynthesis pathway; second
CC step.
CC SIMILARITY: BELONGS TO THE DXR FAMILY.
CC -----
CC his SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC diffied and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



|    |  |  |
|----|--|--|
| RA | Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.U.,  |  |
| RA | Nelson W.C., Gwyn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  |  |
| RA | Halt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  |  |
| RA | Mason T.H., Clacko A., Parksey D.S., Blair E., Citron E., Clark E.B.,  |  |
| RA | Cetoni M.D., Ulterback T.R., Knout H., Qin H., Yamachyan U.,   |  |
| RA | Gill U., Scariato V., Masignani V., Pizzi M., Grandi G., Sun L.,   |  |
| RA | Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,   |  |
| RT | "Complete genome sequence of <i>Neisseria meningitidis</i> serogroup B strain  |  |
| RT | MC58."   |  |
| RL | Science 287.1809-1815(2000).   |  |
| CC | -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction   |  |
| CC | of 1-deoxy-D-xylylose-5-phosphate (DXP) to 2-C-methyl-D-erythritol   |  |
| CC | 4-phosphate (MEP) (By similarity).   |  |
| CC | -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) =  |  |
| CC | 1-deoxy-D-xylylose 5-phosphate + NADPH.  |  |
| CC | -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second  |  |
| CC | step.  |  |
| CC | -1- SIMILARITY: BELONGS TO THE DXR FAMILY.   |  |
| CC | -----  |  |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration   |  |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |  |
| CC | the European Bioinformatics Institute. There are no restrictions on its  |  |
| CC | use by non-profit institutions as long as its content is in no way   |  |
| CC | modified and this statement is not removed. Usage by and for commercial  |  |
| CC | entities requires a license agreement (see <a href="http://www.ldb-sib.ch/announce/">http://www.ldb-sib.ch/announce/</a> |  |
| CC | or send an email to <a href="mailto:license@sib.ch">license@sib.ch</a> ).  |  |
| CC | -----  |  |
| DR | EMBL; AE002375; AAP0641.1; -   |  |
| DR | PIR; A61229; A61229.   |  |
| DR | TIGR; NMB0184;   |  |
| DR | HMAP; MF_00183; -; 1.  |  |
| DR | InterPro; IPR003821; DXR_redicommase.  |  |
| DR | PIfam; PF02670; DXR_reductoisom; 1.  |  |
| DR | TIGRFAMs; TIGR00243; Dxr; 1.   |  |
| KM | Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  |  |
| FT | NP BIND 9 NADPH (POTENTIAL).   |  |
| SC | SEQUENCE 394 AA; 41920 MW; AA853975B5B7BD28 CXC64;   |  |
|    | Query Match 52.4%; Score 1040.5; DB 1; Length 394;   |  |
|    | Best Local Similarity 55.3%; Pred. No. 3e-67;  |  |
|    | Matches 215; Conservative 47; Mismatches 126; Indels 1; Gaps 1;  |  |
| QY | 4 ICIIGATSGISGVSTLDVAVRHHPDKQVVALTANGNIDALTYOCLAHHPREYAVVMESKVA 63   |  |
| DB | 6 LTIIGSTGISGVSTLDVAVSRHPKTRFVALGHQVKEVLAACQOTFPEYAVVADAEHAA 65  |  |
| QY | 64 EFKORIAASPVADIKVYLSGSEALQCVATLEWDFVMAALVGNAGLLPTLAAAKAGKTVIL 123  |  |
| DB | 66 RLEALTKRQDGR - IQVILGAGQLVDVNAADRYSGVMGILVAGVGLPSALAAAGKKTIVL 124   |  |
| QY | 124 ANKEALVNSGQIFWQAVSDSGAVLPIDISEHNATIQCKPAGTYPTHTAKQARRILLTSSG 183   |  |
| DB | 125 ANKETLVVSGALFMETARANGAAVLPVDSHNVAQVLPFDVAGRLNETHIASILLTSSG 184   |  |
| QY | 184 GPFRTPTLEITSSVPPDQAVAPKPKDGRKTSYDSATYMNKKLEIIEACLFFNMEDPDI 243   |  |
| DB | 185 GPFLTADLATPTGRITPDAQVAKPKPKRGRKTSYDSATYMNKKLEIIEHMLFNCPDYLE 244  |  |
| QY | 244 VVIHQSIHSVVDVYDGSVLAQWGNPNPKRTPIAHAMMPEPFDSGVAFLDIFVEGMDP 303  |  |
| DB | 245 VVIHQSIHSVVRDGSVLAQLGNPNPKRTPIACLTGIPERIDSGVGLDPAISALTLE 304   |  |
| QY | 304 EKPDIAKPPICRLAYEATIKSGGIMPTVLAANEIIVEAFINEEVEFTPIAVIIESSMOF 363  |  |
| DB | 305 QEPDIDRPPICRLAYEAMNGGALPCVLAANEAAVAAFLDQIIFTDIAKTVAGCLAD 364   |  |
| QY | 364 KPDDAGSLDELVAQADQDAEYARDIKTL 392   |  |
| DB | 365 FSDGIGDIGILLAQDAPTRAGARAFITGL 393  |  |

AC >XR SALTY STANDARD; PRT; 398 AA.  
AD >B2R23;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 8-FEB-2003 (Rel. 41, Last sequence update)  
DT 8-FEB-2003 (Rel. 41, Last annotation update)  
DE >deoxy-D-xylose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP  
DE >reductoisomerase) (1-deoxyxylulose-5-phosphate reductoisomerase).  
GN OR STM0220.  
OS >Salmonella typhimurium.  
OC >Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC >Enterobacteriaceae; Salmonella.  
OX >ICBI\_taxid=602;  
RN 11  
RP >SEQUENCE FROM N.A.  
RC >TRAIN=lr2 / SCSC1412 / ATCC 700720;  
RX >JDLINR-21534948; Pubmed=11677609;  
RA >McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA >Bourrey L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,  
RA >Leonard S., Nguyen C., Scott K., Holmes A., Grewel N., Molyneux E.,  
RA >Yan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA >Waterston R., Wilson R.K.;  
RT >Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
RT >12";  
RL >ature 413:852-856(2001).  
CC >FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction  
CC >of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol  
CC >4-phosphate (MEP) (By similarity).  
CC >1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)  
CC >= 1-deoxy-D-xylulose 5-phosphate + NADPH.  
CC >1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second  
CC >step.  
CC >1- SIMILARITY: BELONGS TO THE DXR FAMILY.  
CC >-----  
CC >THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC >between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC >at the European Bioinformatics Institute. There are no restrictions on  
CC >use by non-profit institutions as long as its content is in no way  
CC >modified and this statement is not removed. Usage by and for commercial  
CC >activities requires a license agreement (See <http://www.isb-sib.ch/announce>  
CC >or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC >-----  
CC >MBL; AEC008705; AAL19184.1; -  
DR >tyGene; SG2727; dxr.  
DR >TMAP; MF 00183; -; 1.  
DR >InterPro; IPR003821; DXP\_reductoisomase.  
DR >fam; PF02670; DXP\_reductoisom; 1.  
DR >ICRFams; TIGR00243; Dxr; 1.  
DR >soprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
KW >BIND 7 14 NADPH (POTENTIAL).  
FT >SEQUENCE 398 AA; 43352 MW; CB720D95CDBF8FC CRC64;  
SQ >-----  
QY > Match 52.1%; Score 1034; DB 1; Length 398;  
Ma: Local Similarity 55.8%; Pred. No. 8.8e-67;  
Ma: Res 223; Conservative 48; Mismatched 105; Indels 24; Gaps 6  
QY >1 MKGICIGAGTSSIVSTLDVVAARPPDKYQVVALTRANGIDALVYQCLAHBEVAVYVNES 60  
DB >1 MKQITLIGSTISIGCSTLDVVRHNHPPDFSRVIALVAGKNVAAAMADQCLFSPRYAAMDTS 60  
QY >61 KVAEFKRIAAEPVADIKVTSGSALDQVATLVENVDTYMAAIVGAGLLPTLAAAKAGT 120  
DB >61 SAEOLIKMLQOHG-SRTEVLISGGQAAAGEMALDEVGHMAAIVGAAIGLLPTLAAIRAKT 119  
QY >121 VLLANKELVWSGGIIFQMAVDSGAVLLPIISEHNAIFQCP-----AGYTP-----GHT 170  
DB >120 ILLANKESLYTCGLPFDVEKSNARLLIPVDSHNAIFQSLPQSTQHNLTGLADLEQNGVT 179  
QY >171 AKQARRILLTAGSGPFRRTPTETLTSVTPDQVAHPRKMDGRKTSVDSATPANKLEILE 230  
DB >180 S-----ILLTSGSGPFRFETPCDLAAMTPDQACHRPYMSMRKTSVDSATPANKLEILE 234  
QY >231 ACLEFNNEPQDIEVYIHQSTIISHMNVQVDSYLAQNGNPMRTPPIAAVMAPEFDSGV 290

DB 235 ARMLFNASARQMEVLIHPOSIVHSWRYQDSVLAQLGSPDKRPIAHMTAMPKRVTSGA 294

QY 231 APDLFEVGHNDPEKPKRPPCLRLAYEAIKSGIMPTVLNANEIYVAFNEVEKFT 350

DB 235 QPLDCKLSALTFSPADYQRYPCKLAMEAFEGQAATTALNANRITVAFLAQRIET 354

QY 351 DIA----VITERSMAQFKPDGASLEIVLQADQDAREVAR 386

DB 355 DIAGINLAVALERMDIQ---EPASVEDVLYQVDAIAREVAR 390

RESULT 11

ID DNR\_SALTI STANDARD; PRT; 398 AA.

AC 0828A6;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).

GN DXR OR STY0243 OR T0221.

OC Salmonella typhi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

OX NCBI\_TaxId=601;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CT18;

RA MEDLINE=21534947; PubMed=11677608; Thomson N.R., Pickard D., Main J., Parikh J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C., Quail M., Rutherford K., Simmonds W., Skelton J., Stevens K., Whitehead S., Barrett B.G.;

RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typh CT18.";

RT Nature 413:848-852(2001).

RT Nature 413:848-852(2001).

RT [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Ty2 / ATCC 700931;

RA MEDLINE=22531367; PubMed=12644504; Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;

RA "Comparative genomics of Salmonella enterica serovar Typh strains Ty2 and CT18.";

RT J. Bacteriol. 185:2330-2337(2003).

CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (by similarity).

CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xyulose 5-phosphate + NADPH.

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second step.

CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: AL627266; CAD08678.1; -

DR EMBL: AEO16834; AA067951.1; -

DR HAMAP: MF\_00183; -1 1

DR InterPro: IPR003821; DXR\_reductoisomase.

DR Pfam: PF02670; DXR\_reductoisom; 1.

DR TIGR00243; Dxr; 1.

KM Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.

FT NP BIND 7 14 NADPH (POTENTIAL). CRC64;

SO SEQUENCE 398 AA; 43328 MM; 89BA6275828C0E05

Q-ry Match 51.7%; Score 1027; DB 1; Length 398;

Be t Local Similarity 55.8%; Pred. No. 2.8e-66;

M: ches 223; Conservative 49; Mismatches 104; Indels 24; Gaps 7;

QY 1 MKGICITGATGSGIVSTLVDVVAARHPDKRYQVVALTANGNDALYEQCLAAHPEYAVVMES 60

DB 1 MKGILITGSGTSGICSTLVDVHNEPDSFRYIALVAGNVNARMAEQCLEPSPRAYVMDTS 60

QY 61 KVAFKQRIASPVADIKVLSSSEALQVATLENVDVTMAIYGAAGLPTLAAKAGKT 120

DB 61 SAEQLKTMLOOHG-SRTEVLSGOQACENAALEVGHVMAIYGAAGLPTLAAIRAGKT 119

QY 121 VLLANKKALVMSQGIEMQAVSDGAVLTPIDSEHNAIFQCM-----AGTTP-----GHT 170

DB 120 ILLANKESLVTCGLFDEKVRKSYARLLPYDSHNAIFQSLPOSIGHNLGYADLEQNGVT 179

QY 171 AKQARRILTAAGGPFRRPTIETLSVTPQAAVHPKMDGKRIQVDSATMNGKLEIE 230

DB 180 S-----ILLTSGSGPFRRPTIETLMCDLAAMTPQACRHPYMSGRKISVDSATMNGKLEIE 234

QY 231 ACILFPNNEPDQIEVTHPOSTIHSMDVYDGSVLAQMGNDMPRIAHMAMPERRISGV 290

DB 235 ARMLFNASARQMEVLIHPOSIVHSWRYQDSVLAQLGSPDKRPIAHMTAMPKRVTSGA 294

QY 291 APDLFEVGHNDPEKPKRPPCLRLAYEAIKSGIMPTVLNANEIYVAFNEVEKFT 350

DB 295 QPLDCKLSALTFSPADYQRYPCKLAMEAFEGQAATTALNANRITVAFLAQRIET 354

QY 351 DIA----VITERSMAQFKPDGASLEIVLQADQDAREVAR 386

DB 355 DIAGINLAVALERMDIQ---ASVDVLYQVDAIAREVAR 390

REST T 12

DXR IBCH STANDARD; PRT; 402 AA.

AC 09KEV8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).

GN DXR OR VC2254.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.

OX NCBI\_TaxId=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=El Tor N16961 / Serotype O1;

RA MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Zmralava M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;

RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";

RT Nature 406:477-483(2000).

CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol 4-phosphate (MEP) (by similarity).

CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+) = 1-deoxy-D-xyulose 5-phosphate + NADPH.

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second step.

CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: AE004297; AAF95398.1; --  
 CC DR PIR: D82099; D82099.  
 CC DR TIGR: VC2254; --  
 CC DR HAMAP: MF\_00183; -; 1.  
 CC DR InterPro: IPR003821; DXP\_reductoisomerase.  
 CC DR Pfam: PF02670; DXP\_reductoisom; 1.  
 CC DR TIGRPFAMS: TIGR00243; Dxr; 1.  
 CC DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
 CC KW NP\_BIND 7 NADPH (POTENTIAL).  
 CC PT SEQUENCE 402 AA; 43684 MW; 725DF0BCDC27CE6; CRC64;  
 CC SQ

Query Match 51.5%; Score 1023; DB 1; Length 402;  
 Best Local Similarity 52.7%; Pred. No. 5.5e-66;  
 Matches 214; Conservative 65; Mismatches 103; Indels 24; Gaps 4;

QY 1 MKGICILGATGSGIVSTLDVVAHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWES 60  
 1 MKKLTIGATGSGIGASTLKVINQNPQPSIVALVAGVAVAKYGLCQGMREYAVMATAS 60  
 DB 61 KVAEKKRIASPVADIKYLSGSEALQOYATLENDVTMAAIVGAAGLPTLAAKAKKT 120  
 61 NASEL-QGLLNQAMATEVLYGSEAKCOVAAADDDVTMAAIVGAAGLPTLAAKAKKR 119  
 DB 121 VILANKREALVMSGQIFMOAVSDGAVLLPIDSEHNALFQCPAGYTPGHTAKQAR----- 175  
 120 VILANKREALVMSGQIFIDVAGSGALFMPVDSHNALFQCLP-----TETQTDGKCD 172  
 DB 176 -----RLILTAGSGPPFRPTIFLTSVTPDQAAVAPKMDGKISVDSATMNGKIEL 228  
 173 LSGHGIDHLLITGSGGPFERYSDLATLDSVTPEQALAHNWSGPKISVDSATMNGKILEY 232  
 DB 229 IEACLENNERPQIEVVIHPOSIIHSMVYVGVLAQMGNDMPETPLAHMAMPERPDS 288  
 223 IEAKMLNTRSQKLVLIHQSVIHSWQYQDGSVIAQIGEDMTPTISYNALPERVTRA 292  
 DB 289 GVAPIIDIEFVGHMDEKEDLKRFPCLRLAYBAIKSGIMPTVLANANEIAYEALNBEVK 348  
 293 GVPALDEFTLQQLTMEVDFARYPTQLQAMDCFGVATTSILNANAVADAFLKRIK 352  
 DB 349 FFDIAVIERMAQKRPDD-----AGSLNVLQADQDAREVARDIK 390  
 353 FFDIALINDQVLKVCATNTQLHCDLSLELDTMARHFAHQVLK 398

DB 353 FFDIALINDQVLKVCATNTQLHCDLSLELDTMARHFAHQVLK 398

RESULT 13  
 DXR\_YERPE STANDARD; PRT; 398 AA.  
 AC Q8Z862;  
 DT 28-FEB-2003 (Rel. 41; Created)  
 DT 28-FEB-2003 (Rel. 41; Last sequence update)  
 DT 28-FEB-2003 (Rel. 41; Last annotation update)  
 DE 1-deoxy-D-xylofucose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP  
 reductoisomerase) (1-deoxyxylyulose-5-phosphate reductoisomerase).  
 GN DXR OR YPD01048 OR Y3131.  
 OS Versinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Versinia.  
 OX NCBI\_TaxID=632;  
 RN (1)  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=CO-92 / Biotar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parham J. B., Wren B. M., Thomson N. R., Tibball R. W., Holden M. T. G.,  
 RA Prentice M. J., Sebatina M., James K. D., Churcher C., Huggall K. L.,  
 RA Baker S., Basham D., Bentley S. D., Brooks K., Cerdeno-Tarraga A. M.,

RA Millingworth T., Cronin A., Davies R. M., Davis P., Dougan G.,  
 RA Atwell T., Hamlin N., Holroyd S., Jagers K., Kariyeh A. V.,  
 RA Lathier S., Moule S., Oyston P. C. F., Quail M., Rutherford K.,  
 RA Lamond S., Skelton J., Stevens K., Whitehead S., Barrill B. G.,  
 RA The genome sequence of *Versinia pestis*, the causative agent of plague.;  
 RA Nature 413:523-527(2001).  
 RN 2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIMS / Biotar Medisvalis;  
 RX MEDLINE=22137863; PubMed=12142430;  
 RA Ang W., Burland V., Plunkett G. III, Boutin A., Mayhew G. F., Liss P.,  
 RA Jena N. T., Rose D. J., Mau B., Zhou S., Schwartz D. C.,  
 RA Shestern J. D., Lindler L. E., Brubaker R. R., Plano G. V.,  
 RA Cravay S. C., McDonough K. A., Niles M. L., Matsu J. S., Blattner F. R.,  
 RA Jerry R. D.,  
 RA The genome sequence of *Versinia pestis* KIM.;  
 RL Bacteriol. 184:4601-4611(2002).  
 CC FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction  
 CC of 1-deoxy-D-xylofucose 5-phosphate (DXP) to 2-C-methyl-D-erythritol  
 CC 4-phosphate (MEP) (By similarity).  
 CC CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)  
 CC = 1-deoxy-D-xylofucose 5-phosphate + NADPH.  
 CC PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second  
 CC step.  
 CC SIMILARITY: BELONGS TO THE DXR FAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: AJ141416; CAC89890.1; --  
 CC DR EBL: AE013914; AAM86681.1; --  
 CC DR IR: AG0128; AG0128.  
 CC DR HAMAP: MF\_00183; -; 1.  
 CC DR InterPro: IPR003821; DXP\_reductoisomerase.  
 CC DR Pfam: PF02670; DXP\_reductoisom; 1.  
 CC DR TIGRPFAMS: TIGR00243; Dxr; 1.  
 CC DR Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
 CC KW NP\_BIND 7 NADPH (POTENTIAL).  
 CC PT SEQUENCE 398 AA; 43115 MW; C9B1FC950165D057; CRC64;  
 CC SQ

Query Match 51.4%; Score 1020; DB 1; Length 398;  
 Best Local Similarity 55.0%; Pred. No. 8.8e-66;  
 Matches 219; Conservative 51; Mismatches 120; Indels 8; Gaps 3;

QY 1 MKGICILGATGSGIVSTLDVVAHPDKYQVVALTANGNIDALYEQCLAHPEYAVVWES 60  
 1 MKKLTIGATGSGIGASTLKVINQNPQPSIVALVAGVAVAKYGLCQGMREYAVMATAS 60  
 DB 61 KVAEKKRIASPVADIKYLSGSEALQOYATLENDVTMAAIVGAAGLPTLAAKAKKT 120  
 61 S-ASLRILALAGSDTEVYSGRTACELALDDVDQYMAIYGAIPSTLALIRAGQ 119  
 DB 121 VILANKREALVMSGQIFMOAVSDGAVLLPIDSEHNALFQCP-----AGTTPGHTAKQA 174  
 120 VILANKREALVMSGQIFMDVYKRSRAQLPIDSEHNALFQSLPERIQRLQGS- SLNENG 178  
 DB 175 RLILTAGSGPPFRPTIFLTSVTPDQAAVAPKMDGKISVDSATMNGKILEYIAC 234  
 173 RLILTAGSGPPFRPTIFLTSVTPDQAAVAPKMDGKISVDSATMNGKILEYIAR 238  
 DB 179 SRLITSGGPFERYSDLATLDSVTPEQALAHNWSGPKISVDSATMNGKILEYIAR 238  
 235 FNEPDDIIVAVIHPOSIIHSMVYVGVLAQMGNDMPETPLAHMAMPERPDSGVA 294  
 DB 239 FNAAGQIEVYLVLFHQSVIHSWVRYHDSIIAQMGTPMRTPIAHAMAYPKRVSGV 298  
 QY 295 IFVYGHDFEKPDLKRPCLRLAYBAIKSGIMPTVLANANEIAYEALNBEVK 354  
 DB 299 FCKVGLATFTTPDYQRYPCIKLAIIDACNAGQAATTALNANEISVMAFLDSKIR 358



QY 355 IERSMAQKPPDAGSLVLYQADQDAEVARDIKTL 392  
 DB 359 INRTVEGLLSEPTSEVEVLVDRKARDVAAQVIATKL 396

RESULT 14  
 DNR\_HAEMIN STANDARD; PRT; 397 AA.  
 AC P44055;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP  
 reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).  
 GN DNR OR H10807;  
 OS Haemophilus influenzae;  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 NCBI\_TaxID=727;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Kd / K120 / ATCC 51907;  
 RX MEDLINE=9530630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shiley R., Liu L.-I., Glodok A., Kelley J.M.,  
 RA Weiman J.T., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utechtack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd";  
 RL Science 263:496-512 (1995).  
 CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction  
 of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol  
 4-phosphate (MEP) (By similarity).  
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)  
 = 1-deoxy-D-xyulose 5-phosphate + NADPH.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second  
 step.  
 CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: U32763; AAC22466.1; -  
 DR PIR: A64014; A64014.  
 DR TIGR: H10807; -  
 DR HAMAP: MF\_00183; -; 1.  
 DR InterPro: IPR003821; DXP\_reductoisomerase.  
 DR Pfam: PF02670; DXP\_reductoisom; 1.  
 DR TIGRFAMs: TIGR00243; Dxr; 1.  
 KM Isoprene biosynthesis; Oxidoreductase; NADP; Complete proteome.  
 FT NP BIND 9 NADPH (POTENTIAL).  
 SO SEQUENCE 397 AA; 43667 MW; A69B24CCGALCEAA CR664;

Query Match 50.7%; Score 1007; DB 1; Length 397;  
 Best Local Similarity 50.6%; Pseq. No. 7,5e-65;  
 Matches 200; Conservative 76; Mismatches 111; Indels 8; Gaps 3;

QY 2 KSEICIGATGSGIVGTLVAVARHPKQVAVALTANGINDATVGGCLAHPEYAVVWMSK 61  
 DB 4 QMIVLISGSGISGKSTLSTIENNPKTHAFALVGRGVNAMPQICIKRPHRALDDVNA 63  
 QY 62 VAEFKRIASPVADIKVLISGSEALQGVATLENVDVTWMAIVGAAGLLPTLAARAGTV 121

DB 64 AKTIREULIHHI-PTEVLAGRRALCELAHPDADQIMASIVGAAGLLPTLSAVKAGRV 122  
 QY 122 LLANKREALVNSGQITFMQVSDSGAVLIPDISSENNATFOCP-----AGTTPHTAKQAR 175  
 DB 123 LLANKESLVTGCGLFTDAVNRYSKLLPVDSENNATFOSLPPRAQEKIGFCP-LSEIGVS 181  
 QY 176 RILLTAGSGPFRRTPIETLSVTPDDVAHAKMDGKRKISVDSATMMNKGLEIEACILF 235  
 DB 182 KIILTGSGPFRRTPIETLPEQVFNITPEQVVAHPNMSMGKISVDSATMMNKGLEIEARWLF 241  
 QY 236 MNEPDQIEVVIHQSIHSNVDYDVSGLAQKGNPDKRTPIAHAMAPPEFDGAVPLDI 295  
 DB 242 NASAEKEVVIHQSIHSNVRVYDGSVITQKGNPDKRTPIAETMAVPHRTFAGVEPLDF 301  
 QY 296 FEVGHMDPEKDLKRPFCALAYEATKSGSIMPTVLAANEIVAEVAFINEVKFTDIAVI 355  
 DB 302 FKIELFTFIPDFNRPNLKLATIDAPAGGYATTANNAEIVAEVAFINQIGFMIAKI 361  
 QY 356 IERSMAQKPPDAGSLVLYQADQDAEVARDIK 390  
 DB 362 NSKTERISPTTIQINDVLEIDAQAREIAKTLR 396

REST T 15  
 DNR\_XMO STANDARD; PRT; 388 AA.  
 AC 39XEP2; O9RI9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase (EC 1.1.1.267) (DXP  
 reductoisomerase) (1-deoxyxyulose-5-phosphate reductoisomerase).  
 GN DXR.  
 OS Zymomonas mobilis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;  
 OC Sphingomonadaceae; Zymomonas.  
 NCBI\_TaxID=542;  
 RN (1)  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RC STRAIN=ATCC 31821 / ZM4 / CP4;  
 RX MEDLINE=20461176; PubMed=11004410;  
 RA Grollie S., Bringer-Meyer S., Sahn H.;  
 RT "Isolation of the dxr gene of Zymomonas mobilis and characterization  
 of the 1-deoxy-D-xyulose 5-phosphate reductoisomerase.";  
 RL FEBS Microbiol. Lett. 191:131-137 (2000).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 31821 / ZM4 / CP4;  
 RA Lee H.J., Kang H.S.;  
 RP Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Catalyzes the NADP-dependent rearrangement and reduction  
 of 1-deoxy-D-xyulose-5-phosphate (DXP) to 2-C-methyl-D-erythritol  
 4-phosphate (MEP).  
 CC -1- CATALYTIC ACTIVITY: 2-C-methyl-D-erythritol 4-phosphate + NADP(+)  
 = 1-deoxy-D-xyulose 5-phosphate + NADPH.  
 CC -1- COFACTOR: REQUIRES A DIVALENT CATION; MAGNESIUM, MANGANESE OR  
 COBALT.  
 CC -1- ENZYME REGULATION: INHIBITED BY ROSMIDOMYCIN.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; second  
 step.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SIMILARITY: BELONGS TO THE DXR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AJ250714; CAB60758.1; -  
 DR EMBL: AF124757; AA029659.1; -;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 29.3684 Seconds  
(without alignments)  
3461.979 Million cell updates/sec

Title: US-09-941-947a-8  
Perfect score: 1985  
Sequence: 1 MKGICILGATSGISGVSTLDV.....LQADQDAEVARVDIKTLVA 394

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phase:\*  
10: SP\_plant:\*  
11: SP\_prodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description               |
|------------|--------|-------------|--------|-------|---------------------------|
| 1          | 1062   | 53.5        | 402    | 16    | Q8DBF5 Q8DBF5 vibrio vuln |
| 2          | 1041.5 | 52.5        | 396    | 16    | Q8EGG9 Q8EGG9 shewanella  |
| 3          | 983    | 49.5        | 382    | 16    | Q8K43 Q8K43 chlorobium    |
| 4          | 870    | 43.8        | 413    | 16    | Q8DK30 Q8DK30 staphylococ |
| 5          | 838.5  | 42.2        | 472    | 10    | Q8SP64 Q8SP64 attemisia a |
| 6          | 827.5  | 41.7        | 473    | 10    | Q8L6C5 Q8L6C5 stevia reba |
| 7          | 818.5  | 41.2        | 497    | 16    | Q8D2G6 Q8D2G6 wigglewort  |
| 8          | 816.5  | 41.1        | 475    | 10    | Q947C3 Q947C3 lycopersico |
| 9          | 810.5  | 40.8        | 473    | 10    | Q8W250 Q8W250 cryza sativ |
| 10         | 807.5  | 40.7        | 473    | 10    | Q9FTN0 Q9FTN0 cryza sativ |
| 11         | 803.5  | 40.5        | 474    | 10    | Q9M4K4 Q9M4K4 catharanthu |
| 12         | 802.5  | 40.4        | 472    | 10    | Q9FXZ7 Q9FXZ7 zea mays (m |
| 13         | 793.5  | 37.1        | 391    | 16    | Q8T146 Q8T146 leptospira  |
| 14         | 729.5  | 36.8        | 488    | 5     | Q96693 Q96693 plasmodium  |
| 15         | 729.5  | 36.8        | 488    | 5     | Q8IKG4 Q8IKG4 plasmodium  |
| 16         | 718    | 36.2        | 386    | 2     | Q9AJD7 Q9AJD7 kitasatospo |

|    |       |      |      |    |                             |
|----|-------|------|------|----|-----------------------------|
| 1  | 672.5 | 33.9 | 396  | 16 | Q8G7Y7 Q8G7Y7 bifidobacte   |
| 2  | 671   | 33.8 | 504  | 16 | Q8PP80 Q8PP80 corynebacte   |
| 3  | 531   | 26.8 | 377  | 16 | Q8EM05 Q8EM05 mycoplasma    |
| 4  | 495   | 24.9 | 204  | 2  | Q8KMY5 Q8KMY5 escherichia   |
| 5  | 297   | 15.0 | 170  | 10 | Q8LHG5 Q8LHG5 forsythia 1   |
| 6  | 260.5 | 13.1 | 115  | 2  | Q9EXH1 Q9EXH1 klebsiella    |
| 7  | 152   | 7.7  | 94   | 2  | Q87046 Q87046 vibrio chol   |
| 8  | 127   | 6.4  | 781  | 16 | Q8DP912 Q8DP912 vibrio vuln |
| 9  | 115.5 | 5.8  | 1564 | 16 | Q8PR45 Q8PR45 corynebacte   |
| 10 | 114   | 5.7  | 522  | 16 | Q9KM31 Q9KM31 vibrio chol   |
| 11 | 111   | 5.6  | 3104 | 2  | Q04846 Q04846 corynebacte   |
| 12 | 110.5 | 5.6  | 767  | 16 | Q92A84 Q92A84 listeria in   |
| 13 | 110.5 | 5.6  | 7525 | 2  | Q9KIRO Q9KIRO streptomyce   |
| 14 | 109.5 | 5.5  | 6146 | 2  | Q9JH5 Q9JH5 agrobacteri     |
| 15 | 108   | 5.4  | 456  | 16 | Q8UC43 Q8UC43 agrobacteri   |
| 16 | 106   | 5.3  | 486  | 16 | Q8NP76 Q8NP76 corynebacte   |
| 17 | 105.5 | 5.3  | 367  | 16 | Q8Y5X9 Q8Y5X9 listeria mo   |
| 18 | 105.5 | 5.3  | 475  | 16 | Q9AB89 Q9AB89 caulobacter   |
| 19 | 105.5 | 5.3  | 2546 | 2  | Q9A130 Q9A130 burkholderia  |
| 20 | 105.5 | 5.3  | 3816 | 2  | Q9KIV3 Q9KIV3 streptomyce   |
| 21 | 105   | 5.3  | 276  | 16 | Q8NN06 Q8NN06 corynebacte   |
| 22 | 105   | 5.3  | 1096 | 16 | Q8PRK7 Q8PRK7 xanthomonas   |
| 23 | 105   | 5.3  | 2352 | 16 | Q9HTR8 Q9HTR8 pseudomonas   |
| 24 | 104.5 | 5.3  | 395  | 16 | Q92J07 Q92J07 rickettsia    |
| 25 | 104.5 | 5.3  | 1163 | 2  | Q56780 Q56780 xanthomonas   |
| 26 | 104.5 | 5.3  | 9507 | 2  | Q9EMH1 Q9EMH1 streptomyce   |
| 27 | 104   | 5.2  | 405  | 16 | Q98AS0 Q98AS0 rhizobium 1   |
| 28 | 104   | 5.2  | 899  | 16 | Q92273 Q92273 rhizobium m   |
| 29 | 103.5 | 5.2  | 316  | 16 | Q8G115 Q8G115 brucella su   |

## ALIGNMENTS

| Result | ID     | Accession   | Query Match | Length | DB ID | Description               |
|--------|--------|-------------|-------------|--------|-------|---------------------------|
| 1      | Q8DBF5 | PRELIMINARY | 53.5%       | 402    | 16    | Q8DBF5 Q8DBF5 vibrio vuln |
| 2      | Q8DBF5 | PRELIMINARY | 52.5%       | 396    | 16    | Q8EGG9 Q8EGG9 shewanella  |
| 3      | Q8DBF5 | PRELIMINARY | 49.5%       | 382    | 16    | Q8K43 Q8K43 chlorobium    |
| 4      | Q8DBF5 | PRELIMINARY | 43.8%       | 413    | 16    | Q8DK30 Q8DK30 staphylococ |
| 5      | Q8DBF5 | PRELIMINARY | 42.2%       | 472    | 10    | Q8SP64 Q8SP64 attemisia a |
| 6      | Q8DBF5 | PRELIMINARY | 41.7%       | 473    | 10    | Q8L6C5 Q8L6C5 stevia reba |
| 7      | Q8DBF5 | PRELIMINARY | 41.2%       | 497    | 16    | Q8D2G6 Q8D2G6 wigglewort  |
| 8      | Q8DBF5 | PRELIMINARY | 41.1%       | 475    | 10    | Q947C3 Q947C3 lycopersico |
| 9      | Q8DBF5 | PRELIMINARY | 40.8%       | 473    | 10    | Q8W250 Q8W250 cryza sativ |
| 10     | Q8DBF5 | PRELIMINARY | 40.7%       | 473    | 10    | Q9FTN0 Q9FTN0 cryza sativ |
| 11     | Q8DBF5 | PRELIMINARY | 40.5%       | 474    | 10    | Q9M4K4 Q9M4K4 catharanthu |
| 12     | Q8DBF5 | PRELIMINARY | 40.4%       | 472    | 10    | Q9FXZ7 Q9FXZ7 zea mays (m |
| 13     | Q8DBF5 | PRELIMINARY | 37.1%       | 391    | 16    | Q8T146 Q8T146 leptospira  |
| 14     | Q8DBF5 | PRELIMINARY | 36.8%       | 488    | 5     | Q96693 Q96693 plasmodium  |
| 15     | Q8DBF5 | PRELIMINARY | 36.8%       | 488    | 5     | Q8IKG4 Q8IKG4 plasmodium  |
| 16     | Q8DBF5 | PRELIMINARY | 36.2%       | 386    | 2     | Q9AJD7 Q9AJD7 kitasatospo |

QY 174 ARIILLTASGGFRRPTLETTSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 233  
 DB 178 VSHILLTGSGGFFRYTVDVAEIAVPEQAIAPENMSMEFKISVDBATMMNKLEIIEAKW 237  
 QY 234 LFNMPDQIEVYIHQSIIHSMWVDVDSVLAQKGNPKRPTIAPAMAPRFDGVAPL 293  
 DB 238 LFNARSDQIKVLIHQSVIHSWQYDGSVLAQKMGSPKATYIALTIPERVAKGACPL 297  
 QY 294 DIFEVGHMDFEKPKDLKRPCLRLAYEAIKSGIMPTVLAANEIAVEAFINEEVPTDIA 353  
 DB 298 DFTQVGLTFLQPDFFRYPCALAIACACYLQGHATTTLTNAANEVAAFLAQLKFTDIA 357  
 QY 354 -----VIERSMAQFKDDAGSELEVYQADQDAEVARDIK 390  
 DB 358 RANSDVLANQVCKQSLASGLDSLELTDPMRTTADDEVYR 398

## RESULT 2

08EGG9 PRELIMINARY; PRT; 396 AA.  
 AC 08EGG9;  
 DT 01-MAR-2003 (TReMBLrel, 23, Created)  
 DT 01-MAR-2003 (TReMBLrel, 23, last sequence update)  
 DT 01-MAR-2003 (TReMBLrel, 23, last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.  
 GN DXR OR S01635.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadales; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.T., Nelson W.C.,  
 Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 Meyer T., Tsaplin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
 DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 Vamathevan J., Weidman J., Impraim M., Lee C., Berry K., Lee C.,  
 Mueller J., Khouri H., Gill J., Uutterback T.R., McDonald L.A.,  
 Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123 (2002).  
 DR EMBL; AB015609; AAN54690.1; -  
 DR TIGR; S01635; -  
 KW Isomerase; Complete proteome.  
 SQ SEQUENCE 396 AA; 42159 MW; EDB1785DDE37A3B7 CRC64;

Query Match 52.5%; Score 1041.5; DB 16; Length 396;  
 Best local similarity 55.1%; Pred. No. 2.3e-69;  
 Matches 217; Conservative 49; Mismatches 121; Indels 7; Gaps 2;

QY 1 MKGICILGATGSGVSTLDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYVMS 60  
 DB 1 KGMVTLGATGSGVSTLDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYVMS 60  
 QY 61 KVAEFKRIAPVADIKVLSGEALQOVALLENVDVTMAAIVGAAGLPTLAAGRT 120  
 DB 61 KVAEFKRIAPVADIKVLSGEALQOVALLENVDVTMAAIVGAAGLPTLAAGRT 120  
 DB 61 KVAEFKRIAPVADIKVLSGEALQOVALLENVDVTMAAIVGAAGLPTLAAGRT 120  
 QY 121 VILANKKALVMSGGIEMQAVSDSGAVLLPIDESENAIFQCPAGTTPG-----TKAKQAR 175  
 DB 121 VILANKKALVMSGGIEMQAVSDSGAVLLPIDESENAIFQCPAGTTPG-----TKAKQAR 175  
 QY 176 RILTFASGPPRRPTLETTSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 235  
 DB 176 RILTFASGPPRRPTLETTSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 235  
 QY 238 LFNARSDQIKVLIHQSVIHSWQYDGSVLAQKMGSPKATYIALTIPERVAKGACPL 297  
 DB 238 LFNARSDQIKVLIHQSVIHSWQYDGSVLAQKMGSPKATYIALTIPERVAKGACPL 297  
 QY 294 DIFEVGHMDFEKPKDLKRPCLRLAYEAIKSGIMPTVLAANEIAVEAFINEEVPTDIA 353  
 DB 294 DIFEVGHMDFEKPKDLKRPCLRLAYEAIKSGIMPTVLAANEIAVEAFINEEVPTDIA 353  
 QY 354 -----VIERSMAQFKDDAGSELEVYQADQDAEVARDIK 390  
 DB 354 -----VIERSMAQFKDDAGSELEVYQADQDAEVARDIK 390  
 QY 358 RANSDVLANQVCKQSLASGLDSLELTDPMRTTADDEVYR 398  
 DB 358 RANSDVLANQVCKQSLASGLDSLELTDPMRTTADDEVYR 398

DB 239 MTQDQIKVLIHQSVIHSWQYDGSVLAQKMGSPKATYIALTIPERVAKGACPL 297  
 QY 296 FEVGHMDFEKPKDLKRPCLRLAYEAIKSGIMPTVLAANEIAVEAFINEEVPTDIA 355  
 DB 299 FKVGQLSFCEPDFRRPCALAIACACQCGQEGEATVLAANEIAVEAFINEEVPTDIA 358  
 QY 356 IERSMAQFKDDAGSELEVYQADQDAEVARDIK 389  
 DB 359 NEDCLTSVPKQAMASIEDITLADQRTIARIEL 392

## REST 1 3

08K 3 PRELIMINARY; PRT; 382 AA.  
 AC 08K43;  
 DT 01-OCT-2002 (TReMBLrel, 22, Created)  
 DT 01-OCT-2002 (TReMBLrel, 22, last sequence update)  
 DT 01-MAR-2003 (TReMBLrel, 23, last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.  
 GN DXR OR C70125.  
 OS Chlorobium tepidum.  
 OC Bacteria; Chlorobi; Chlorobiales; Chlorobiaceae;  
 OC Chlorobium.  
 OX NCBI\_TaxID=1097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TL5 / ATCC 49652 / DSM 12025;  
 RX MEDLINE=22103685; PubMed=12093901;  
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
 Dodson R.J., DeBoy R., Gwinn M.L., Nelson W.C., Haft D.H.,  
 Hickey B.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
 Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
 Niernman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,  
 Vamathevan J., Khouri H., Bryant D.A., Fraser C.M.,  
 Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.,  
 RT "The complete genome sequence of Chlorobium tepidum TL5, a  
 RT photosynthetic, anaerobic, green-sulfur bacterium.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).  
 DR EMBL; AB012791; AAM71373.1; -  
 DR TIGR; C70125; -  
 DR InterPro; IPR003821; DXR reductoisomerase.  
 DR Pfam; PF02670; DXR reductoisom; 1.  
 DR TIGRFAMs; TIGR00243; DXR; 1.  
 KW Isomerase; Complete proteome.  
 SQ SEQUENCE 382 AA; 41343 MW; 4B428B6F36368CA CRC64;

Query Match 49.5%; Score 983; DB 16; Length 382;  
 Best local similarity 50.6%; Pred. No. 5e-65;  
 Mismatch 198; Conservative 75; Mismatches 106; Indels 12; Gaps 3;

QY 1 MKGICILGATGSGVSTLDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYVMS 60  
 DB 1 MKGICILGATGSGVSTLDVVAHPPKQYVALTANGNIDALYEOCLAHPEYAVYVMS 60  
 QY 61 KVAEFKRIAPVADIKVLSGEALQOVALLENVDVTMAAIVGAAGLPTLAAGRT 118  
 DB 61 KVAEFKRIAPVADIKVLSGEALQOVALLENVDVTMAAIVGAAGLPTLAAGRT 118  
 QY 119 KTVILANKKALVMSGGIEMQAVSDSGAVLLPIDESENAIFQCPAGTTPGHTAKQAR 178  
 DB 119 KTVILANKKALVMSGGIEMQAVSDSGAVLLPIDESENAIFQCPAGTTPGHTAKQAR 178  
 DB 119 KTVILANKKALVMSGGIEMQAVSDSGAVLLPIDESENAIFQCPAGTTPGHTAKQAR 178  
 QY 179 LTFASGPPRRPTLETTSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 238  
 DB 179 LTFASGPPRRPTLETTSSVTPDQAVAHPKMDGKISVDSATMANKLEIIEACL 238  
 QY 238 LFNARSDQIKVLIHQSVIHSWQYDGSVLAQKMGSPKATYIALTIPERVAKGACPL 297  
 DB 238 LFNARSDQIKVLIHQSVIHSWQYDGSVLAQKMGSPKATYIALTIPERVAKGACPL 297  
 QY 294 DIFEVGHMDFEKPKDLKRPCLRLAYEAIKSGIMPTVLAANEIAVEAFINEEVPTDIA 353  
 DB 294 DIFEVGHMDFEKPKDLKRPCLRLAYEAIKSGIMPTVLAANEIAVEAFINEEVPTDIA 353  
 QY 354 -----VIERSMAQFKDDAGSELEVYQADQDAEVARDIK 390  
 DB 354 -----VIERSMAQFKDDAGSELEVYQADQDAEVARDIK 390  
 QY 358 RANSDVLANQVCKQSLASGLDSLELTDPMRTTADDEVYR 398  
 DB 358 RANSDVLANQVCKQSLASGLDSLELTDPMRTTADDEVYR 398

QY 359 SNAQFKPDDAGSLBLVLAQADODAREVARDI 389  
 DB 351 TMOAHEAWPTLBRVLAQADKWARBTAQOLI 381

RESULT 4  
 Q8DK30 PRELIMINARY; PRT; 413 AA.

AC Q8DK30;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.  
 GN TLR1040.  
 OS Synchococcus elongatus (Thermosynchococcus elongatus).  
 OC Bacteria; Cyanobacteria; Chromococcales; Synchococcus.  
 OK NCBI\_Taxid=32046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BP-1;  
 RX MEDLINE=2225144; PubMed=12240834;  
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
 RA Katsumura A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the thermophilic cyanobacterium  
 RT Thermosynchococcus elongatus BP-1.";  
 RL EMBL, AF005372; BAC08593.1;  
 DR DNA Res. 9:123-130(2002).  
 KW Isomerase; Complete proteome.  
 SQ SEQUENCE 413 AA; 44148 MW; 7276081507960644 CRC64;

Query Match 43.8%; Score 870; DB 16; Length 413;  
 Best local Similarity 45.4%; Pred. No. 1.5e-56;  
 Matches 184; Conservative 76; Mismatches 123; Indels 22; Gaps 6;

QY 1 MGICILGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHHPYAVVWMSK 60  
 DB 19 VQALNLTSGTSGTGLDVAQYDRFRVGLAAGNLERIPQIRQFQPIVSIADPE 78  
 QY 61 KVAEFQRILASPVDADIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAARAGKT 120  
 DB 79 QLPETLALADLPORP-QVAGEAGIAAAYAGDBVVTGIVGAGVPTIATKAGD 137  
 QY 121 VLLANKALWMSGQFMQAVSDSGAVLPIIDSEHNAIQCM---PAGYTPGHTAQARI 177  
 DB 138 IALANKETLIAGGPVLPILQYGVKLPADSEHSAIFQCLQGVPG-----GIKRI 189  
 QY 178 LITASGGPFRRTPIETLSVTPDQAVAHPKMDGSKISVDSATMNNKGLIELIACLLFNM 237  
 DB 190 ILTASGAFRDPVPELQAVTADALKEPNMMSGPRITVDSATLNKGLIEVIAHYLFQM 249  
 QY 238 EPDQIEVVIHPQSIHSMVDYDGSVLAQMGNDPRTPIAHMANPERPDSVAALDIFE 297  
 DB 250 DYDNLEIVHPQSIHSLIEDTSLAQLGMPDRLPLVLTLSWPERITMSPDLVAK 309  
 QY 298 VGHMDPEKEDLRFPCRLAVERAISKGGIMPTVLANAEIAVEAFLEVEKFTDIAVIER 357  
 DB 310 ADDLTFRSDHQKYPGMGLAYAAAGAGAMPVLANAAGVALFIAELIATLEIPRIIE 369  
 QY 358 ----RSMQFRTDDAGSLBLVLAQADODAR---EVARDIIXLIVA 394  
 DB 370 WVCDRYSQNIINP--TLEDIIAADRMARATVOEALQRGVSQVVA 412

RESULT 5  
 Q8SP64 PRELIMINARY; PRT; 472 AA.

AC Q8SP64;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.  
 GN X1.  
 OS Artemisia annua (Sweet wormwood).  
 OC Caryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Ceratophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Steridae; campanulids; Asterales; Asteraceae; Asteroideae;  
 OC Anthemideae; Artemisia.  
 OK NCBI\_Taxid=35608;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Yugoslavica; TISSUE=Root;  
 RA Debe K.K., Scout F.F., Shore K.A., Weathers P.J.;  
 RT Artemisia annua 1-deoxy-D-xyulose-5-phosphate reductoisomerase (dxr)  
 GN "A.";  
 RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 RL XBL; AF182287; AAD56391.2;  
 DR InterPro; IPR003821; DXP reductoisomerase.  
 DR Fam; PF02670; DXP reductoisom; 1.  
 DR GRFAMS; TIGR00243; Dxr; 1.  
 KW Isomerase.  
 SQ SEQUENCE 472 AA; 50740 MW; D52023C09D475675 CRC64;

Query Match 42.2%; Score 838.5; DB 10; Length 472;  
 Best local Similarity 46.2%; Pred. No. 3.9e-54;  
 Matches 181; Conservative 66; Mismatches 134; Indels 11; Gaps 5;

QY 2 KGIICILGATGSGVSTLDVVAARHPDKYQVVALTANGNIDALYEQCLAHHPYAVVWMSK 61  
 DB 75 KPIIITGSGTSGTGLDVAENPDKRVALLAGSVTLTAEQIKAFKPOLVSIKNESL 134  
 QY 62 VAEFRQRIASPVDADIKVLSGSEALQOVATLENVDYMAAIVGAAGLLPTLAARAGKT 121  
 DB 135 VAEIKBALAGSDVMP-ELIPDEGVVAVARHPDCVTVVTGIVGAGLKPVTAAIEAGKNI 193  
 QY 122 LLAANKALWMSGQFMQAVSDSGAVLPIIDSEHNAIQCM---PAGYTPGHTAQARI 181  
 DB 194 ALANKETLIAGGPVLPILAHKRNKILPADSEHSAIFQCLQGVPG-----ALKRI 248  
 QY 182 SGCPFRRTPIETLSVTPDQAVAHPKMDGSKISVDSATMNNKGLIELIACLLFNM 241  
 DB 249 SGCAFDPMPVKLADVAVADALKEPNMMSGKILVDSATLNKGLIEVIAHYLGSSYDN 308  
 QY 242 IEVVIHPQSIHSMVDYDGSVLAQMGNDPRTPIAHMANPERPDSVAP---LDIFEV 298  
 DB 309 IDIVHPQSIHSMVDYDGSVLAQMGNDPRTPIAHMANPERPDSVAP---LDIFEV 368  
 QY 299 GHMDPEKEDLRFPCRLAVERAISKGGIMPTVLANAEIAVEAFLEVEKFTDIAVIER 358  
 DB 369 GSLTFKADNVKYSMSHLAYSAGAGGTGCVLSAANEKAVEMTLDKIGTLDFKVEL 428  
 QY 359 SNAQFKPD--DAGSLBLVLAQADODAREVARDI 388  
 DB 429 TCEHQALVLTAPLSEIHHYDLARREYAAV 460

RESULT 6  
 Q8E6C PRELIMINARY; PRT; 473 AA.

AC Q8E6C;  
 DT 1-OCT-2002 (TREMBlrel. 22, Created)  
 DT 1-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 1-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.  
 GN X1.  
 OS Artemisia annua (Sweet wormwood).  
 OC Caryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Ceratophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Steridae; campanulids; Asterales; Asteraceae; Asteroideae;  
 OC Anthemideae; Artemisia.  
 OK NCBI\_Taxid=35608;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N.M.L.C.;  
 RT Artemisia annua 1-deoxy-D-xyulose-5-phosphate reductoisomerase (dxr)  
 GN "A.";  
 RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 RL XBL; AF182287; AAD56391.2;  
 DR InterPro; IPR003821; DXP reductoisomerase.  
 DR Fam; PF02670; DXP reductoisom; 1.  
 DR GRFAMS; TIGR00243; Dxr; 1.  
 KW Isomerase.  
 SQ SEQUENCE 473 AA; 50740 MW; D52023C09D475675 CRC64;

RT "Cloning and expression of cDNAs encoding two enzymes of the MEK  
 pathway in *Stevia rebaudiana* Bertoni.",  
 RL Plant Physiol. 0:0-0(0).  
 DR EMBL: A0429233; CAD22156.1; -.  
 DR InterPro: IPR003821; DXP\_redismase.  
 DR Pfam: PF02670; DXP\_redismase; 1.  
 DR TrEMBL: TIGR00243; DXP; 1.  
 KW isomerase.  
 SQ SEQUENCE 473 AA; 51021 MW; 03BD5BF204B62490 CRC64;

QY 175 REILLTASGPRRTPIETLSVTDPQVAHHPKMKERKISVDSATMMKIELIEACLL 234  
 DB 246 REILLTASGPRRTPIETLSVTDPQVAHHPKMKERKISVDSATMMKIELIEACLL 305  
 QY 235 FMEPDQIEVVIHPOSIIHSMVDYDGSVLAQMGNDKRTPIAHAMPER-FDSGVA-- 291  
 DB 306 FCAEYDNIIEIVIHPOSIIHSMVETQDSVLAQMGNDKRTPIAHAMPER-FDSGVA-- 365  
 QY 292 FLDIEVGHMDKPRDLKRPCLRLAYEAIKSGGIMPTVLAANEIAYEAFLEBEVKTLD 351  
 DB 366 FLDIEVGHMDKPRDLKRPCLRLAYEAIKSGGIMPTVLAANEIAYEAFLEBEVKTLD 425  
 QY 352 IAVIIEKSAQKPRD--DAGSELVLAQMGNDKRTPIAHAMPER-FDSGVA-- 388  
 DB 426 IPRIVELTCAHREELVSPSLIEIITHDUMARDYAASTL 464

RESULT 9  
 ID 08W250 PRELIMINARY; PRT; 473 AA.  
 AC 08W250;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase precursor  
 (EC 1.1.1.-).  
 OS Oryza sativa (Rice).  
 OC Burkholderia: Vitridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Burharoidae; Oryzaceae; Oryza.  
 NCBI\_TaxID=4530;  
 (1)  
 RA SEQUENCE FROM N.A.  
 RA Caretero-Paulet L., Boronat A., Campos N.;  
 RA 11-deoxy-D-xyulose 5-phosphate reductoisomerase (DXR), catalyzing the  
 RT first committed step of the mevalonate-independent pathway for IPP  
 RT biosynthesis.  
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF367205; AAL37560.1; -.  
 DR Gramene; 08W250; -.  
 DR InterPro; IPR003821; DXP reductoisomerase.  
 DR Pfam; PF02670; DXP reductoisom; 1.  
 DR TIGRFAMs; TIGR00243; Dxr; 1.  
 KW Transit peptide; Isomerase; Oxidoreductase.  
 FT TRANSIT 1  
 FT POTENTIAL.  
 SQ SEQUENCE 473 AA; 51497 MW; 618C4A675283478F CRC64;

Query Match 40.8%; Score 810.5; DB 10; Length 473;  
 Best Local Similarity 45.4%; Pred. No. 48e-52;  
 Matches 182; Conservative 67; Mismatches 123; Indels 29; Gaps 8;

QY 2 KGCITLIGATGSGTLDVVAHHPKMKERKISVDSATMMKIELIEACLL 61  
 DB 77 KPISTIVSTGSGTLDVVAHHPKMKERKISVDSATMMKIELIEACLL 136  
 QY 62 VAEFKQRIASPVADIK--VLSGSEALQOVATLENTVTVAIYGAAGLPTLAAKAGK 119  
 DB 137 VDELKEALAD---CDWKEPILPGEQVTEVAHHPKMKERKISVDSATMMKIELIEACLL 193  
 QY 120 TVLANKKALVMSQIFMQAVSDSGAVLLPDSSENAIFOC---MPAGTTPGHAKQARR 176  
 DB 194 DIALANKKELIAGSPFVLPLAKKRVKILPADSEHSAIFOCIOGLPEG-----ALRR 245  
 QY 177 ILLTASGPRRTPIETLSVTDPQVAHHPKMKERKISVDSATMMKIELIEACLL 236  
 DB 246 ILLTASGPRRTPIETLSVTDPQVAHHPKMKERKISVDSATMMKIELIEACLL 305  
 QY 237 MEPPDQIEVVIHPOSIIHSMVDYDGSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 293  
 DB 306 AVEYDNIIEIVIHPOSIIHSMVETQDSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 365  
 QY 294 DIFEVGHMDKPRDLKRPCLRLAYEAIKSGGIMPTVLAANEIAYEAFLEBEVKTLD 353

DB 366 DLCKGSLTFRKADPNVYKPSMDLAAAGRAGTMTGVLAAANEKAYELFIDEKIGYDIF 425  
 QY 354 VIE-----RSMAGKPRDAGSELVLAQMGNDKRTPIAHAMPER-FDSGVA-- 388  
 DB 426 KVELTCDAHNELVTSR-----SLEIITHDUMAREVAASTL 464

RESULT 10  
 ID 08FTN0 PRELIMINARY; PRT; 473 AA.  
 AC 08FTN0;  
 DT 1-MAR-2001 (TREMBLrel. 16, Created)  
 DT 1-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 1-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE 1-deoxy-D-xyulose 5-phosphate reductoisomerase.  
 OS Oryza sativa (Rice).  
 OC Burkholderia: Vitridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Burharoidae; Oryzaceae; Oryza.  
 NCBI\_TaxID=4530;  
 (1)  
 RA SEQUENCE FROM N.A.  
 RA TRAIN-cv. Nipponbare;  
 RA Asaki T., Matsumoto T., Yamamoto K.;  
 RA Oryza sativa nipponbare (GAJ) genomic DNA, chromosome 1, PAC  
 RT clone: P0482C06.2;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF002845; BAB16915.1; -.  
 DR Gramene; 08FTN0; -.  
 DR InterPro; IPR003821; DXP reductoisomerase.  
 DR Pfam; PF02670; DXP reductoisom; 1.  
 DR TIGRFAMs; TIGR00243; Dxr; 1.  
 KW Isomerase.  
 SQ SEQUENCE 473 AA; 51473 MW; 9933D9D8C3D1FC49 CRC64;

Query Match 40.7%; Score 807.5; DB 10; Length 473;  
 Best Local Similarity 45.4%; Pred. No. 8e-52;  
 Matches 182; Conservative 66; Mismatches 124; Indels 29; Gaps 8;

QY 2 KGCITLIGATGSGTLDVVAHHPKMKERKISVDSATMMKIELIEACLL 61  
 DB 77 KPISTIVSTGSGTLDVVAHHPKMKERKISVDSATMMKIELIEACLL 136  
 QY 62 VAEFKQRIASPVADIK--VLSGSEALQOVATLENTVTVAIYGAAGLPTLAAKAGK 119  
 DB 137 VDELKEALAD---CDWKEPILPGEQVTEVAHHPKMKERKISVDSATMMKIELIEACLL 193  
 QY 120 TVLANKKALVMSQIFMQAVSDSGAVLLPDSSENAIFOC---MPAGTTPGHAKQARR 176  
 DB 194 DIALANKKELIAGSPFVLPLAKKRVKILPADSEHSAIFOCIOGLPEG-----ALRR 245  
 QY 177 ILLTASGPRRTPIETLSVTDPQVAHHPKMKERKISVDSATMMKIELIEACLL 236  
 DB 246 ILLTASGPRRTPIETLSVTDPQVAHHPKMKERKISVDSATMMKIELIEACLL 305  
 QY 237 MEPPDQIEVVIHPOSIIHSMVDYDGSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 293  
 DB 306 AVEYDNIIEIVIHPOSIIHSMVETQDSVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 365  
 QY 294 DIFEVGHMDKPRDLKRPCLRLAYEAIKSGGIMPTVLAANEIAYEAFLEBEVKTLD 353  
 DB 366 DLCKGSLTFRKADPNVYKPSMDLAAAGRAGTMTGVLAAANEKAYELFIDEKIGYDIF 425  
 QY 354 VIE-----RSMAGKPRDAGSELVLAQMGNDKRTPIAHAMPER-FDSGVA--PL 388

Db 426 KVELTCDARHNEVTRP---SLEETIHDLMAREYASL 462

## RESULT 11

OSM4M4 PRELIMINARY; PRT; 474 AA.

AC 09FXM4: 01-OCT-2000 (TREMELrel. 15, Created)  
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)  
 DE 1-deoxy-D-xylulose-5-phosphate reductoisomerase.  
 DXR.  
 OS Catharanthus roseus (Rose periwinkle) (Madagascar periwinkle).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; Lamiales; Gentianales; Apocynaceae; Rauvolfiaceae; Vinaceae;  
 OC Catharanthus.  
 OC NCBI\_TaxID=4058;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20568781; PubMed=11116631;  
 RA Veau B., Courtois M., Oudin A., Chénieux J.-C., Rideau X., Claeste M.;  
 RT "Cloning and expression of cDNAs encoding two enzymes of the MEP  
 pathway in Catharanthus roseus."  
 RL Biochim. Biophys. Acta 1517:159-163(2000).  
 DR EMBL; AF50235; AAF65154.1; -  
 DR InterPro: IPR003821; DXP\_reductoisomase.  
 DR Pfam: PF02670; DXP\_reductoisom; 1.  
 DR TIGRFAMs; TIGR00243; Dxr; 1.  
 DR Isomerase.  
 SQ SEQUENCE 474 AA; 51165 MW; C2737CD318B0C94C CRC64;

Query Match 40.5%; Score 803.5; DB 10; Length 474;  
 Best Local Similarity 44.0%; Pred. No. 1.6e-51;  
 Matches 176; Conservative 71; Mismatches 134; Indels 19; Gaps 6;

QY 2 KGIICLGATGSGIGVSTLDVVAARHPDKYQVVALTAMGNIDALVEQCLAHPEYAVVWESK 61  
 DB KPISTIVGSGVGTOTLIDVAENPDKFRVVALAGSNVTLADQVTFKPKLVAVRNESL 137  
 QY 62 VAEFKQRIASPVAD--IKYLSGSEALQVATLENDVTMAALVGAAGILPTLAAGAKXT 120  
 DB 138 VNSLSE--ALSDVDDPELITPGQGVVEVRHSDAVTVGTGVGAGLKPVALAAGKD 195  
 QY 121 VILANKKALVMSGOIFMQAVSDSGAVLLPIDSEHNAIFQCC--MDAGYTPGHTAQAARI 177  
 DB 196 IALANKETILAGKPPVLLPLAHKHKVILPDADEHSAIFQCIQGLPEG-----ALRRI 247  
 QY 178 LITAGSGPRRTPIETLSSTVTPDQAVAHPRKMDGKISVDSATYMMKGLIEACLIETRM 237  
 DB 248 ILTASGGAARDMPVEELKGVKVAADALKHPNMWGGKIVDSATLTFNKGLEVEIAHYLFQA 307  
 QY 238 EPPDLEFVTHPOSIHSMVDYDGSVLAQMGNDPKRTPIAHAMWPERFDSGVA--LD 294  
 DB 308 EVDNDIVIHPOSIHSMVETODSSVLAQLGMPDMRLPLVLTLSMPDRISCEIIMPRLD 367  
 QY 295 IREVGMDPEKPDLPKPCRLALAYBAIKSGGIMPTVLANAEIAYEAPLNEVEKTTDIAV 354  
 DB 368 LCKLGSILFTKTPDNVYKPSMDLAAAGAGGTGTGTLASANKKAVELPIDEKISTYLDIFK 427  
 QY 355 ITERSMAOKPD--DAGSLELVLAQDQDARVARDIKTL 392  
 DB 428 VVELTGNARHNELVTPSLDEIIVHDLGARDYAASTQNSL 467

## RESULT 12

09FXM4 PRELIMINARY; PRT; 472 AA.

AC 09FXM4: 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

DE Putative 1-deoxy-D-xylulose 5-phosphate reductoisomerase.  
 GN DXR.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 OC NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=dwair-1; TISSUE=Root;  
 RA Hains J., Hains B., Strack D., Walter M.H.;  
 RT "Regulation of the non-mevalonate methylerythritol phosphate (MEP)  
 pathway by mycorrhizal fungi."  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ297566; CAC03581.1; -  
 DR InterPro: IPR003821; DXP\_reductoisomase.  
 DR Pfam: PF02670; DXP\_reductoisom; 1.  
 DR TIGRFAMs; TIGR00243; Dxr; 1.  
 DR Isomerase.  
 SQ SEQUENCE 472 AA; 51285 MW; 9D5E30F8C8A0706C CRC64;

Query Match 40.4%; Score 802.5; DB 10; Length 472;  
 Best Local Similarity 45.1%; Pred. No. 1.9e-51;  
 Matches 178; Conservative 68; Mismatches 132; Indels 17; Gaps 6;

QY 2 KGIICLGATGSGIGVSTLDVVAARHPDKYQVVALTAMGNIDALVEQCLAHPEYAVVWESK 61  
 DB 76 KPISTIVGSGVGTOTLIDVAENPDKFRVVALAGSNVTLADQVTFKPKLVAVRNESL 135  
 QY 62 VAEFKQRIASPVAD--IKYLSGSEALQVATLENDVTMAALVGAAGILPTLAAGAKXT 118  
 DB 136 VDELKRALACERKP---ELIPGQGVIEVAHPDVAVTGTGVGAGLKPVALAAG 191  
 QY 119 KTVLANKEALVMSGOIFMQAVSDSGAVLLPIDSEHNAIFQCC--MDAGYTPGHTAQAARI 178  
 DB 192 KDIANKKALVMSGOIFMQAVSDSGAVLLPIDSEHNAIFQCC--MDAGYTPGHTAQAARI 246  
 QY 179 LITAGSGPRRTPIETLSSTVTPDQAVAHPRKMDGKISVDSATYMMKGLIEACLIETRM 238  
 DB 247 ILTASGGAARDMPVEELKGVKVAADALKHPNMWGGKIVDSATLTFNKGLEVEIAHYLFQA 306  
 QY 239 EPPDLEFVTHPOSIHSMVDYDGSVLAQMGNDPKRTPIAHAMWPERFDSGVA--LD 295  
 DB 307 YDDIEIYTHPOSIHSMVETODSSVLAQLGMPDMRLPLVLTLSMPDRISCEIIMPRLD 366  
 QY 296 IREVGMDPEKPDLPKPCRLALAYBAIKSGGIMPTVLANAEIAYEAPLNEVEKTTDIAV 355  
 DB 367 CKLGSILFTKTPDNVYKPSMDLAAAGAGGTGTGTLASANKKAVELPIDEKISTYLDIFK 426  
 QY 356 ITERSMAOKPD--DAGSLELVLAQDQDARVARDIKTL 388  
 DB 427 VVELTGNARHNELVTPSLDEIIVHDLGARDYAASTQNSL 461

## RESULT 13

09FXM4 PRELIMINARY; PRT; 391 AA.

AC 09FXM4: 01-MAR-2003 (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE 1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.-).  
 GN DXR OR LA3292.  
 OS Leptosiphra interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OC NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;  
 RA Ren S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF011490; AAN50490.1; -  
 DR Isomerase; Oxidoreductase; Complete proteome.



| QY      | SEQUENCE   | 391 AA;      | 43026 MM;          | 8CF56BC0C17AC0F7 CRC64; |
|---------|--|--------------|--------------------|-------------------------|
| QY      | Query Match  | 37.1%;       | Score 735.5;       | DB 16; Length 391;      |
| Db      | Best Local Similarity  | 41.0%;       | Pred. No. 1.4e-46; |                         |
| Matches | 158; Conservative  | 78;          | Mismatches 140;    | Indels 9; Gaps 4        |
| QY      | 4 ICILGATGSGVSTLDVVAHPDKYQVVALTANGNIDALYEQCLAHPEVAVVMESKVA         | 63           |                    |                         |
| Db      | 7 VCLLGASGSGVSTLTKVLAHPDKRIRHFSVHSLERAKLQKEFSDF--ICVSSDFA          | 64           |                    |                         |
| QY      | 64 EEPKRIASBPVADIKVLSSSEALQOVALLENVDTMAALVGAAGLLPTLAARAKTALL       | 123          |                    |                         |
| Db      | 65 DV--GVLAENLGRTOILYESSLCEIYEPPEVEIYITLVSGVGRPTIAITFGKTLGI        | 122          |                    |                         |
| QY      | 124 ANKEALVMSQIMQAVSDSGAVLLPISEHNAIFQCPAGTQGHAKARRILLTASG          | 183          |                    |                         |
| Db      | 123 ANKEVLTVSPPQLQSLIAKNTKVPVDSHNAALQTLLES--LNPVSVEK---ILLTASG     | 177          |                    |                         |
| QY      | 184 GPRFRPTPIETLSSVTPDQVAHPKQDMGRKLSVDSATMMNKGLEIELNACLEFNNPEPOIB  | 243          |                    |                         |
| Db      | 178 GAFRRDLPVEQLSVSTKEQLHPHTMNGPKRTIDNSGINKGLFVLEAFLEFVNPYDKIG     | 237          |                    |                         |
| QY      | 244 VVHHPQSIHSMVDYDVGSLAQMGKPNKRTPIALAMAPFERPQSGVAPLDIPEVGMDF      | 303          |                    |                         |
| Db      | 238 VVHHPQSIAGVIELKDGASFLYASYPMIPEIASLHPPEPVKRVLSRYPAKMGKLEF       | 297          |                    |                         |
| QY      | 304 EKPDLPKPCILRAYEALIKSGIMPTLVNAIEALVEAFLEEVKFTDIAVILERSMAQF      | 363          |                    |                         |
| Db      | 298 REEDFPKRPCLGIAFLAGKVGGTAPCIRFANAEAAVELFLDHRIFLEIPYIRETIDEI     | 357          |                    |                         |
| QY      | 364 KPDDAGSLBELVLAQDDAREVARDI                                      | 388          |                    |                         |
| Db      | 358 KIEFPPLSLEVEADRIARETVRNLL                                      | 382          |                    |                         |
| QY      | RESULT 14  |              |                    |                         |
| QY      | 096693   |              |                    |                         |
| ID      | 096693   | PRELIMINARY; | PRT;               | 488 AA.                 |
| AC      | 096693;  |              |                    |                         |
| DT      | 01-MAY-1999 (TEMBREIrel. 10, Created)                              |              |                    |                         |
| DT      | 01-MAY-1999 (TEMBREIrel. 10, Last sequence update)                 |              |                    |                         |
| DT      | 01-JUN-2002 (TEMBREIrel. 21, Last annotation update)               |              |                    |                         |
| DE      | 1-deoxy-D-xyulose 5-phosphate reductoisomerase.                    |              |                    |                         |
| GN      | DNR.   |              |                    |                         |
| OS      | Plasmodium falciparum.   |              |                    |                         |
| OC      | Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.       |              |                    |                         |
| OX      | NCBI_Taxid=5831;   |              |                    |                         |
| RN      | [1]  |              |                    |                         |
| RP      | SEQUENCE FROM N.A.   |              |                    |                         |
| RC      | STRAIN=HB3;  |              |                    |                         |
| RA      | Altinicek B., Sanderbrand S., Wiesner J., Jomaa H.;                |              |                    |                         |
| RT      | "dxr as a potential target for antimalarial drugs.";               |              |                    |                         |
| RL      | Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.            |              |                    |                         |
| DR      | EMBL; AF118113; AAD03739.1; "                                      |              |                    |                         |
| DR      | InterPro: IPR003821; DXP reductoisomase.                           |              |                    |                         |
| DR      | Pfam: PF02670; DXP reductoisom; 1.                                 |              |                    |                         |
| DR      | TIGRFAMs; TIGR00243; Dxr; 1.                                       |              |                    |                         |
| KM      | Isomerase.   |              |                    |                         |
| SO      | Isomerase. 488 AA; 55756 MM; 4E280C81CFAD3EF CRC64;                |              |                    |                         |
| QY      | Query Match  | 36.8%;       | Score 729.5;       | DB 5; Length 488;       |
| Db      | Best Local Similarity  | 37.8%;       | Pred. No. 5.4e-46; |                         |
| Matches | 153; Conservative  | 86;          | Mismatches 141;    | Indels 23; Gaps 7       |
| QY      | 4 ICILGATGSGVSTLDVVAHPDK---YQVVALTANGNIDALYEQCLAHPEVAVVMESKVA      | 59           |                    |                         |
| Db      | 80 VAFEGTGTSGTMAANI--RECKKLENFVNVVALYVNSVNELYEQARELPFELCIHDR       | 138          |                    |                         |
| QY      | 60 SKVAEPKRIASBPVADIK--VLSSSEALQOVALLENVDTMAALVGAAGLLPTLAARAK      | 117          |                    |                         |
| Db      | 139 SVYEELEKELV--KNIDYRPILIGDDECKEIKCSNSIDKIVIGIDSGQSYSTMVAILN     | 166          |                    |                         |
| QY      | 118 GKTVLLANKREALVMSQIMQAVS--DSGAVLLPISEHNAIFQCPAGTQGHAKARRILLTASG | 173          |                    |                         |

|         |  |              |  |         |
|---------|--|--------------|--|---------|
| Dd      |  | 197          | NKTVLANNESIYSAQFLAKLNLNHRKAIIIPDSSHSALFQCLDNKKYLKTKCLODNF      | 256     |
| Oy      |  | 174          | -----ARRILTLASGGPFRPFRIETLTSSVTPDOAVAHPRKMGRKXISVDATMONTGLEL   | 228     |
| Dd      |  | 257          | SKINNINXIKFLCSSGGGFONLTMDCLKNVTSENALKHPKMTMGKKIRIDBATMMANKLELV | 316     |
| Oy      |  | 229          | IEACLLFNNEPPOIEIVYHPOSIIHSWPDYDVGYLAQMGPDMRPTIAHAMAPRFOS       | 288     |
| Dd      |  | 317          | LETHFLFPDYNDIEVYHKECTIIHSCVEPIFKSVISGMYPDOQLITLYSLWPDRKIT      | 376     |
| Oy      |  | 289          | GVAPLDIPEVGHGMDPEKDPDKRPCLRLAYAIKSGGIMPTVLNAANEIYAFALEEVK      | 348     |
| Dd      |  | 377          | NLPEDLDAQVSTLTTRHKPSLEHFPCITKLAYAGAKGNFYPLVANANSEIANMLFLNNKIK  | 436     |
| Oy      |  | 349          | FTTIVATIERSMAGKRPDDAGS-----LELVADODQAREFARPI                   | 388     |
| Dd      |  | 437          | YFOISSIIISQVLIESFNSOKSENSIEDLMKQILLHSMAXKDXTDI                 | 481     |
| RESULT: | 15   |              |  |         |
| GOING   |  |              |  |         |
| ID      | 3IXG4  | PRELIMINARY; | PRT;   | 488 AA. |
| AC      | 3IXG4  |              |  |         |
| DT      | 1-MAR-2003 (TREMBLrel_23, Created)                                     |              |  |         |
| DT      | 1-MAR-2003 (TREMBLrel_23, Last sequence update)                        |              |  |         |
| DT      | MAR-2003 (TREMBLrel_23, Last annotation update)                        |              |  |         |
| DE      | deoxy-D-xylulose 5-phosphate reductoisomerase.                         |              |  |         |
| CN      | 114_0641.  |              |  |         |
| OC      | Isomodum falciparum (isolate 3D7).                                     |              |  |         |
| CS      | Karyote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.             |              |  |         |
| OX      | BI_TaxID=36329;  |              |  |         |
| RN      |  |              |  |         |
| RP      | SOURCE FROM N.A.   |              |  |         |
| RC      | RAIN=3D7;  |              |  |         |
| FX      | 3DITMB=2255705; PubMed=12368864;                                       |              |  |         |
| RA      | Irdner M.J., Hall N., Fung E., White C., Berriman M., Hyman R.M.,      |              |  |         |
| RA      | Trilhon U.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K., |              |  |         |
| RA      | Asen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,            |              |  |         |
| RA      | Jean M.-S., Neue V., Shalton S.J., Sub B., Peterson J., Anguoli S.,    |              |  |         |
| RA      | Berthe M., Allen J., Selengut J., Hatz D., Mather M.W., Valdaye A.B.,  |              |  |         |
| RA      | Arltin D.M.A., Fairlamb A.H., Franholz M.J., Roos D.S., Ralph S.A.,    |              |  |         |
| RA      | Radgen G.I., Cummings L.M., Subramanian G.M., Mungall C.,              |              |  |         |
| RA      | Inter J.C., Garucci D.J., Hoffman S.L., Newbold C., Davis R.W.,        |              |  |         |
| RA      | asser C.M., Bartell B.;  |              |  |         |
| RT      | Genome sequence of the human malaria parasite Plasmodium               |              |  |         |
| RT      | falciparum."   |              |  |         |
| RL      | Structure 419:498-511(2002).   |              |  |         |
| DR      | CEL; AB014826; AANJ37254.1; -.   |              |  |         |
| KM      | Cometase 488 AA; 55756 MW; 4E280C81CDFADF3EF CRC64;                    |              |  |         |
| SO      | SOURCE   |              |  |         |
| Match   | 36.8%; Score 729.5; DB 5; Length 488;                                  |              |  |         |
| Bes:    | Local Similarity 37.8%; Pred. No. 5.4e-46;                             |              |  |         |
| Mat:    | ies 153; Conservative 88; Mismatches 141; Indels 23; Gaps 7            |              |  |         |
| Oy      |  | 4            | ICLLGTGSIGYSTLDVVAHRDK-----YGVVLTANONDAILEOGLAHHREYAAVYME      | 59      |
| Dd      |  | 80           | VAFPGSTSIGTNALNTI-RBCNKLENFVNKALYVANSVNELTEQARFREPTELCHDK      | 138     |
| Oy      |  | 60           | SKTAERFORIAASPVAADIK--VLSSGEALOOVATLENVTMAIVAAGILPTLAAKA       | 117     |
| Dd      |  | 139          | SYVEELKEHY--KNIKDYKPIITLCGDDEMKRICSGNSIDIKVIDBSFGYGISTMYAKM    | 196     |
| Oy      |  | 118          | GKVYLLANKEALWMSGQFMQAVS--DSQAVLLPIDSEHALFQQMPAGYTPGHTAQ--      | 173     |
| Dd      |  | 197          | NKIVALANESIVSASFPLKLIHLNHAKKIIIPDSSHSAIFQCLDNKKYLKTKCLODNF     | 256     |
| Oy      |  | 174          | -----ARRILTLASGGPFRPFRIETLSVTPDOAVAHPRKMGRKXISVDATMONTGLEL     | 228     |
| Dd      |  | 257          | SKINNINXIKFLCSSGGGFONLTMDCLKNVTSENALKHPKMTMGKKIRIDBATMMANKLELV | 316     |
| Oy      |  | 229          | IEACLLFNNEPPOIEIVYHPOSIIHSWPDYDVGYLAQMGPDMRPTIAHAMAPRFOS       | 288     |
| Dd      |  | 317          | LETHFLFPDYNDIEVYHKECTIIHSCVEPIFKSVISGMYPDOQLITLYSLWPDRKIT      | 376     |
| Oy      |  | 289          | GVAPLDIPEVGHGMDPEKDPDKRPCLRLAYAIKSGGIMPTVLNAANEIYAFALEEVK      | 348     |
| Dd      |  | 377          | NLPEDLDAQVSTLTTRHKPSLEHFPCITKLAYAGAKGNFYPLVANANSEIANMLFLNNKIK  | 436     |
| Oy      |  | 349          | FTTIVATIERSMAGKRPDDAGS-----LELVADODQAREFARPI                   | 388     |
| Dd      |  | 437          | YFOISSIIISQVLIESFNSOKSENSIEDLMKQILLHSMAXKDXTDI                 | 481     |
| RESULT: | 15   |              |  |         |
| GOING   |  |              |  |         |
| ID      | 3IXG4  | PRELIMINARY; | PRT;   | 488 AA. |
| AC      | 3IXG4  |              |  |         |
| DT      | 1-MAR-2003 (TREMBLrel_23, Created)                                     |              |  |         |
| DT      | 1-MAR-2003 (TREMBLrel_23, Last sequence update)                        |              |  |         |
| DT      | MAR-2003 (TREMBLrel_23, Last annotation update)                        |              |  |         |
| DE      | deoxy-D-xylulose 5-phosphate reductoisomerase.                         |              |  |         |
| CN      | 114_0641.  |              |  |         |
| OC      | Isomodum falciparum (isolate 3D7).                                     |              |  |         |
| CS      | Karyote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.             |              |  |         |
| OX      | BI_TaxID=36329;  |              |  |         |
| RN      |  |              |  |         |
| RP      | SOURCE FROM N.A.   |              |  |         |
| RC      | RAIN=3D7;  |              |  |         |
| FX      | 3DITMB=2255705; PubMed=12368864;                                       |              |  |         |
| RA      | Irdner M.J., Hall N., Fung E., White C., Berriman M., Hyman R.M.,      |              |  |         |
| RA      | Trilhon U.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K., |              |  |         |
| RA      | Asen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,            |              |  |         |
| RA      | Jean M.-S., Neue V., Shalton S.J., Sub B., Peterson J., Anguoli S.,    |              |  |         |
| RA      | Berthe M., Allen J., Selengut J., Hatz D., Mather M.W., Valdaye A.B.,  |              |  |         |
| RA      | Arltin D.M.A., Fairlamb A.H., Franholz M.J., Roos D.S., Ralph S.A.,    |              |  |         |
| RA      | Radgen G.I., Cummings L.M., Subramanian G.M., Mungall C.,              |              |  |         |
| RA      | Inter J.C., Garucci D.J., Hoffman S.L., Newbold C., Davis R.W.,        |              |  |         |
| RA      | asser C.M., Bartell B.;  |              |  |         |
| RT      | Genome sequence of the human malaria parasite Plasmodium               |              |  |         |
| RT      |  |              |  |         |

```

Db      317 IETHPLFDVNDYNDIEVIVHKECIIHSCVEFIDKSVISOMYTPDMOQIPILYSLTWPDRIKT 376
Qy      289 GVAPLDIEFVGMDPEKPDILKRFPCILRLAYEAIKSGGIMPTVLNANEIAYEAPLINEVK 348
Db      377 NLKPLDLAQVSTLTFPHKPSLEHFPCTILAYQAGIKGNFYPTVLNANSNEIANNLFINNKIK 436
Qy      349 FTDIAVIERSMQKFPDDAGS-----LEVLQADQDAREVARDI 388
Db      437 YFDISSIISQVLESENSQKVSSENSDLMKQILQIHSWAKDKATDI 481

```

Search completed: January 29, 2004, 15:54:04  
 Job time : 31.3684 secs

```

APPLICANT: Wang, Man
APPLICANT: Mooney, Jeffrey L.
APPLICANT: Debonck, Christine M.
APPLICANT: Zhong, Yi Yi
TITLE OF INVENTION: No. 6204042el GimU
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Dechart Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Street
City: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PATSCO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/309,026
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/971,782
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10024
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 459 amino acids
TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
/US-09-309-026-2

```

|                       |        |                 |       |                |
|-----------------------|--------|-----------------|-------|----------------|
| Query Match           | 7.4%;  | Score 88.5;     | DB 3; | Length 459;    |
| Best Local Similarity | 23.9%; | Pred. No. 0.12; |       |                |
| Matches               | 48;    | Conservative    | 32;   | Mismatches 74; |
|                       |        |                 |       | Indels 47;     |
|                       |        |                 |       | Gaps 9;        |

```

QY      MAVVPAGGKRRQAPRQCVPLGKQVITHTTLRLLESAPQKVAIVSVEPWPBEL 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      4 FALILAGGTRRKSLPVTYHKVAGISMLH-----VFSVC-AIOPE----- 46

QY      68 SIARHPDIIITAPGKERADSVLSALKALEDIASENDVYLVDHA--ARPCLTGSDHLQI 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      47 -----KTVLVGHR--AEIVBEVLAGQEPETQSEQGTGHAVMHTPELLEGSHTIV 98

QY      125 DTLRNDPVGILALSSHDTLKVAGDGTITATIDRKHVRAITPOM--FKYGM-----R 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      99 -----IAGDPPLITGESLKNL-----IDFHNHKVAITLTAEIDNPFGRIVRNDNA 147

QY      177 DALQRTGKNPAVTDEASALEL 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      148 EVLRMVEQDADDFEQKIKSI 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: January 29, 2004, 15:57:04  
Job time : 8.2499 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 ; Search time 15.7685 Seconds  
(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947a-10

Perfect score: 1199  
Sequence: 1 MNPPTICMAVVPAGVGKRM.....IKITRPEDLALQPTMEQA 231

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 769580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 769580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEM\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                           |
|------------|-------|-------------|--------|-------|---------------------------------------|
| 1          | 1199  | 100.0       | 231    | 10    | US-09-934-903-6 Sequence 68, Appl     |
| 2          | 1199  | 100.0       | 231    | 10    | US-09-934-868-6 Sequence 68, Appl     |
| 3          | 1199  | 100.0       | 231    | 11    | US-09-941-947a-10 Sequence 10, Appl   |
| 4          | 555.5 | 46.3        | 234    | 9     | US-09-815-242-11931 Sequence 11931, A |
| 5          | 555.5 | 44.2        | 225    | 9     | US-09-815-242-11076 Sequence 11076, A |
| 6          | 529.5 | 44.2        | 254    | 9     | US-09-815-242-11722 Sequence 11722, A |
| 7          | 520.5 | 43.4        | 236    | 9     | US-09-815-242-10273 Sequence 10273, A |
| 8          | 520.5 | 43.4        | 236    | 10    | US-09-797-251-11 Sequence 11, Appl    |
| 9          | 499.5 | 41.7        | 236    | 9     | US-09-815-242-13867 Sequence 13867, A |
| 10         | 297   | 24.8        | 232    | 10    | US-09-792-251-8 Sequence 8, Appl      |
| 11         | 283.5 | 23.6        | 302    | 10    | US-09-935-943-8 Sequence 8, Appl      |
| 12         | 255.5 | 22.3        | 332    | 12    | US-10-128-713A-6 Sequence 6, Appl     |
| 13         | 245.5 | 20.5        | 250    | 15    | US-10-156-761-11503 Sequence 11503, A |
| 14         | 240.5 | 20.1        | 218    | 12    | US-10-289-762-621 Sequence 621, App   |
| 15         | 223.5 | 18.6        | 256    | 10    | US-09-738-626-6422 Sequence 6422, Ap  |

|   |       |      |      |    |                     |                   |
|---|-------|------|------|----|---------------------|-------------------|
| 1 | 217.5 | 18.1 | 228  | 9  | US-09-815-242-12471 | Sequence 12471, A |
| 1 | 214.5 | 17.9 | 227  | 9  | US-09-815-242-5711  | Sequence 5711, Ap |
| 1 | 214.5 | 17.9 | 238  | 9  | US-09-816-181-2     | Sequence 2, Appl  |
| 1 | 209   | 17.4 | 235  | 9  | US-09-815-242-13256 | Sequence 13256, A |
| 2 | 209   | 17.4 | 235  | 10 | US-09-792-251-2     | Sequence 2, Appl  |
| 2 | 198.5 | 16.6 | 238  | 10 | US-09-816-182-2     | Sequence 2, Appl  |
| 2 | 193.5 | 16.1 | 236  | 9  | US-09-815-242-5707  | Sequence 5707, Ap |
| 2 | 193.5 | 16.1 | 242  | 9  | US-09-815-242-12467 | Sequence 12467, A |
| 2 | 193.5 | 16.1 | 242  | 9  | US-09-815-242-12467 | Sequence 12467, A |
| 2 | 105.5 | 8.8  | 291  | 12 | US-10-369-493-21638 | Sequence 21638, A |
| 2 | 98.5  | 8.2  | 461  | 9  | US-09-815-242-10937 | Sequence 10937, A |
| 2 | 98.5  | 8.2  | 783  | 10 | US-09-738-626-5017  | Sequence 18528, A |
| 2 | 98.5  | 8.2  | 395  | 12 | US-10-369-493-18528 | Sequence 20306, A |
| 2 | 94    | 7.8  | 327  | 12 | US-10-369-493-20306 | Sequence 17444, A |
| 3 | 93    | 7.8  | 297  | 12 | US-10-369-493-17444 | Sequence 245, App |
| 3 | 93    | 7.8  | 357  | 12 | US-10-289-762-487   | Sequence 487, App |
| 3 | 91    | 7.6  | 419  | 12 | US-10-369-493-21686 | Sequence 21686, A |
| 3 | 90.5  | 7.5  | 285  | 12 | US-10-369-493-1153  | Sequence 1153, Ap |
| 3 | 90.5  | 7.5  | 456  | 9  | US-09-815-242-10400 | Sequence 10400, A |
| 3 | 89.5  | 7.5  | 381  | 16 | US-10-080-170-264   | Sequence 264, App |
| 3 | 89.5  | 7.5  | 385  | 12 | US-10-369-493-21535 | Sequence 21535, A |
| 3 | 89.5  | 7.5  | 154  | 9  | US-09-939-980-528   | Sequence 528, App |
| 3 | 89    | 7.4  | 704  | 15 | US-10-128-714-8238  | Sequence 8238, Ap |
| 4 | 89    | 7.4  | 1317 | 12 | US-10-369-493-13693 | Sequence 13693, A |
| 4 | 88.5  | 7.4  | 361  | 12 | US-10-369-493-23443 | Sequence 23443, A |
| 4 | 88.5  | 7.4  | 485  | 10 | US-09-738-626-4551  | Sequence 4551, Ap |
| 4 | 88    | 7.3  | 394  | 12 | US-10-369-493-11824 | Sequence 11824, A |
| 4 | 87.5  | 7.3  | 443  | 15 | US-10-156-761-14528 | Sequence 14528, A |
| 4 | 87.5  | 7.3  | 831  | 15 | US-10-156-761-14505 | Sequence 14505, A |

#### ALIGNMENTS

RESULT 1  
US-09-934-903-6  
Sequence 6, Application US/09934903  
Pat. No. US20020102690A1  
GENE2AL INFORMATION:  
APPLICANT: Kofkas, Mathios  
APPLICANT: Odom, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. US20020102690A1ton, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Rouviere, Pierre  
APPLICANT: Platagisio, Stephen  
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
PRIOR REFERENCE: C11646 US NA  
CURRENT APPLICATION NUMBER: US/09/934,903  
CURRENT FILING DATE: 2001-08-22  
PCT APPLICATION NUMBER: 60/229,907  
PCT FILING DATE: September 1, 2001  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 6  
LENGTH: 231  
TAX: PRT  
ORGANISM: Methylobacter 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ORF3  
US-09-934-903-6  
Query / Match 100.0%; Score 1199; DB 10; Length 231;  
Best Local Similarity 100.0%; Pred. No. 2.1e-115;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 MNPPTICMAVVPAGVGKRMADPRKYLPLAGTIVHTLTRESDAPKVAIVSV 60  
DB 1 MNPPTICMAVVPAGVGKRMADPRKYLPLAGTIVHTLTRESDAPKVAIVSV 60  
CY 61 DPYBPRIAKHPDITAPGCKERADSVLSALRALIEDIASNDVVLVDAPPCITGSDI 120

Db 61 DPMPELSIAKHEDITTAAGKERADSVLSALKALEDIASENDVWVHDAARPCLTGSDI 120  
Qy 121 HLOIDTLKNDPVGGIILASHSDTLKGVGDITTAITIDRKHWRALTPOMFKYGLRDALQ 180  
Db 121 HLOIDTLKNDPVGGIILASHSDTLKGVGDITTAITIDRKHWRALTPOMFKYGLRDALQ 180  
Qy 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231  
Db 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231

## RESULT 2

US-09-934-868-68  
Sequence 68, Application US/09934868  
Patent No. US20020137190A1  
GENERAL INFORMATION:  
APPLICANT: Koffas, Mattheos  
APPLICANT: Odum, James M.  
APPLICANT: Schenzle, Andreas J.  
TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN  
FILE REFERENCE: CL1596 US NA  
CURRENT APPLICATION NUMBER: US/09/934,868  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229,858  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 68  
LENGTH: 231  
TYPE: PRT  
ORGANISM: Methylomonas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ISPD  
US-09-934-868-68

Query Match 100.0%; Score 1199; DB 10; Length 231;  
Best Local Similarity 100.0%; Pred. No. 2,1e-115;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MNPITQCAVAVPAAGVGRKQADRPQYLPAGKTYIEHTLTLLSSDAFOKVAVAISVE 60  
Db 1 MNPITQCAVAVPAAGVGRKQADRPQYLPAGKTYIEHTLTLLSSDAFOKVAVAISVE 60  
Qy 61 DPMPELSIAKHEDITTAAGKERADSVLSALKALEDIASENDVWVHDAARPCLTGSDI 120  
Db 61 DPMPELSIAKHEDITTAAGKERADSVLSALKALEDIASENDVWVHDAARPCLTGSDI 120  
Qy 121 HLOIDTLKNDPVGGIILASHSDTLKGVGDITTAITIDRKHWRALTPOMFKYGLRDALQ 180  
Db 121 HLOIDTLKNDPVGGIILASHSDTLKGVGDITTAITIDRKHWRALTPOMFKYGLRDALQ 180  
Qy 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231  
Db 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231

## RESULT 3

US-09-941-947A-10  
Sequence 10, Application US/09941947A  
Publication No. US20030003528A1  
GENERAL INFORMATION:  
APPLICANT: Brzostowicz, Patricia C.  
APPLICANT: Cheng, Qiong  
APPLICANT: DiCosimo, Deana J.  
APPLICANT: Koffas, Mattheos  
APPLICANT: Miller, Edward S. Jr.  
APPLICANT: Odum, J. Martin  
APPLICANT: Picataggio, Steve  
APPLICANT: Roviello, Pierre F.  
TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE  
FILE REFERENCE: CL1903 US NA  
CURRENT APPLICATION NUMBER: US/09/941,947A

CURRENT FILING DATE: 2001-09-01  
FOR APPLICATION NUMBER: 60/229,907  
FOR FILING DATE: 2000-09-01  
FOR APPLICATION NUMBER: 60/229,858  
FOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 10  
LENGTH: 231  
TYPE: PRT  
ORGANISM: Methylomonas 16a  
US-09-941-947A-10

Query Match 100.0%; Score 1199; DB 11; Length 231;  
Best Local Similarity 100.0%; Pred. No. 2,1e-115;  
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MNPITQCAVAVPAAGVGRKQADRPQYLPAGKTYIEHTLTLLSSDAFOKVAVAISVE 60  
Db 1 MNPITQCAVAVPAAGVGRKQADRPQYLPAGKTYIEHTLTLLSSDAFOKVAVAISVE 60  
Qy 61 DPMPELSIAKHEDITTAAGKERADSVLSALKALEDIASENDVWVHDAARPCLTGSDI 120  
Db 61 DPMPELSIAKHEDITTAAGKERADSVLSALKALEDIASENDVWVHDAARPCLTGSDI 120  
Qy 121 HLOIDTLKNDPVGGIILASHSDTLKGVGDITTAITIDRKHWRALTPOMFKYGLRDALQ 180  
Db 121 HLOIDTLKNDPVGGIILASHSDTLKGVGDITTAITIDRKHWRALTPOMFKYGLRDALQ 180  
Qy 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231  
Db 181 RTEGNPAVTDEASALELIGHKPKIVEGRPDNIKITRPEDLALAQFYMEQQA 231

## RESULT 4

US-09-815-242-11931  
Sequence 11931, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Orlsen, Karl L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
THE OF INVENTION: Prokaryotes  
FILE REFERENCE: BL17A, 011a  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
FOR APPLICATION NUMBER: 60/191,078  
FOR FILING DATE: 2000-03-21  
FOR APPLICATION NUMBER: 60/206,848  
FOR FILING DATE: 2000-05-23  
FOR APPLICATION NUMBER: 60/207,727  
FOR FILING DATE: 2000-05-26  
FOR APPLICATION NUMBER: 60/242,578  
FOR FILING DATE: 2000-10-23  
FOR APPLICATION NUMBER: 60/253,625  
FOR FILING DATE: 2000-11-27  
FOR APPLICATION NUMBER: 60/257,931  
FOR FILING DATE: 2000-12-32  
FOR APPLICATION NUMBER: 60/269,308  
FOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11931  
LENGTH: 234  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-11931

Query Match 46.3%; Score 555.5; DB 9; Length 234;  
 Best Local Similarity 54.1%; Pred. No. 5.3e-49;  
 Matches 119; Conservative 26; Mismatches 72; Indels 3; Gaps 3;

QY 8 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLRLIESDAFQKVAVAIVEDPYWPEL 67  
 DB 10 WTVIPAAVGSRRADRRKQYLDLAGRTVIERLDCFLHEHMLRGLVYCLAEDDPYWPGL 69  
 QY 68 SIAGKPDITTAAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDT 126  
 DB 70 DCAARHYQRRAGGKERAGSVLNGLRILELGAQDDVAVHDAARPCLTGSDIHLIEE 129  
 QY 127 LKNDVGGIILASHTLKHVDG-TTATIDRKHWRAALTPQMKYGMRLDAL-QRTG 184  
 DB 130 LAEDVGGIILAVPADTLKESDRDGRVSEITIDRSVVMATYPPQWRLGALRALDALVA 189  
 QY 185 NPATDEASALELGHKPKIVGREDNKTIRPEDLALAQ 224  
 DB 190 GVAITDEASALEAGVAPRLVGRADNKTIRPEDLRLQ 229

RESULT 5  
 US-09-815-242-11076  
 ; Sequence 11076, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: EITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 11076  
 ; LENGTH: 225  
 ; TYPE: PRF  
 ; ORGANISM: Haemophilus influenzae  
 ; US-09-815-242-11076

Query Match 44.2%; Score 530.5; DB 9; Length 225;  
 Best Local Similarity 52.5%; Pred. No. 1.9e-46;  
 Matches 117; Conservative 28; Mismatches 65; Indels 13; Gaps 6;

QY 9 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLRLIESDAFQKVAVAIVEDPYWPEL 68  
 DB 7 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLRLIESDAFQKVAVAIVEDPYWPEL 66  
 QY 69 IAKHPDITTAAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDTLK 128  
 DB 67 L--DEKIDVGGITTAAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDTLK 117

QY 129 --NDPVGGIILASHTLKHVDG-TTATIDRKHWRAALTPQMKYGMRLDALQ--TEG 184  
 DB 118 AIBKQALIALPVTDLTKRADNGQCIKTEBDSQLMQAMTPOGFPDILRDLSTGIDQ 177  
 QY 185 NPATDEASALELGHKPKIVGREDNKTIRPEDLALAQFYM 227  
 DB 178 GANTDEASALELAGFPRLVGRADNKTIRPEDLALAEFYL 220

RESULT 6  
 US-09-815-242-11722  
 ; Sequence 11722, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Karl L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: EITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 11722  
 ; LENGTH: 254  
 ; TYPE: PRF  
 ; ORGANISM: Klebsiella pneumoniae  
 ; US-09-815-242-11722

Query Match 44.2%; Score 529.5; DB 9; Length 254;  
 Best Local Similarity 51.1%; Pred. No. 2.9e-46;  
 Matches 114; Conservative 30; Mismatches 70; Indels 9; Gaps 5;

QY 9 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLRLIESDAFQKVAVAIVEDPYWPEL 68  
 DB 33 AAVPAAGVGRMQRADPKQYPLAGKTVIEHTLRLIESDAFQKVAVAIVEDPYWPEL 92  
 QY 69 IAKHPDITTAAGKERADSVSA-LKALEDIASENDVAVHDAARPCLTGSDIHLQIDTLK 128  
 DB 93 LAHPQITTVVGGIILASHTLKHVDG-TTATIDRKHWRAALTPQMKYGMRLDALQ--TEG 148  
 QY 129 NDPVGGIILASHTLKHVDG-TTATIDRKHWRAALTPQMKYGMRLDALQ--TEG 184  
 DB 149 TSVGGIILAVPADTLKESDRDGRVSEITIDRSVVMATYPPQWRLGALRALDALVA 208  
 QY 185 NPATDEASALELGHKPKIVGREDNKTIRPEDLALAQFYM 227  
 DB 209 AATIDDEASALELAGFPRLVGRADNKTIRPEDLALAEFYL 250

RESULT 7  
 US-09-815-242-10273

```

; Sequence 10273, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10273
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-815-242-10273

Query Match      43.4%; Score 520.5; DB 9; Length 236;
Best Local Similarity 49.8%; Pred. No. 2.2e-45;
Matches 111; Conservative 31; Mismatches 72; Indels 9; Gaps 4;

QY 9 AVPAAGVGKMGQADRPKQYLPAGKTIVIEHTLTRLESADAFQKVAIVSVDPYMELS 68
DB 10 AVPAAGVGKMGQADRPKQYLPAGKTIVIEHTLTRLESADAFQKVAIVSVDPYMELS 69
QY 69 IAKHPDITTAAGKGRADSVLSALKLEDIASENDVVLVHDARPCLTGSDIHLQIDTLK 128
DB 70 LANHPQITVVDGDERADSVLAGLKA---AGDAQWVLVHDARPCLTGSDIHLQIDTLK 125
QY 129 NDVPGILALSHDITLKHVD--GDTTATIDRKHWVRLTPQWKKYGMULDALQ--TEG 184
DB 126 TSRTGGILAPVDTMKRABPGKNAIAHYVDKNGMLHALLTPQFFPRELLHDCITRALNEG 185
QY 185 NPATVDEASALELIGKRPKIVGSRPNIKITRPEDLALAQFYM 227
DB 186 -ATTDEASALEYCGFHPQLVGRADNIKVTREPDALAEFYL 227

RESULT 8
US-09-792-251-11
; Sequence 11, Application US/09792251
; Patent No. US20020160364A1
; GENERAL INFORMATION:
; APPLICANT: Fritzt, Christian
; APPLICANT: Youngman, Philip
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: USE OF YACM AND YOEJ ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 06286-140001
; CURRENT APPLICATION NUMBER: US/09/792,251
; NUMBER OF SEQ ID NOS: 28

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-792-251-11

Query Match      43.4%; Score 520.5; DB 10; Length 236;
Best Local Similarity 49.8%; Pred. No. 2.2e-45;
Matches 111; Conservative 31; Mismatches 72; Indels 9; Gaps 4;

QY 9 AVPAAGVGKMGQADRPKQYLPAGKTIVIEHTLTRLESADAFQKVAIVSVDPYMELS 68
DB 10 AVPAAGVGKMGQADRPKQYLPAGKTIVIEHTLTRLESADAFQKVAIVSVDPYMELS 69
QY 69 IAKHPDITTAAGKGRADSVLSALKLEDIASENDVVLVHDARPCLTGSDIHLQIDTLK 128
DB 70 LANHPQITVVDGDERADSVLAGLKA---AGDAQWVLVHDARPCLTGSDIHLQIDTLK 125
QY 129 NDVPGILALSHDITLKHVD--GDTTATIDRKHWVRLTPQWKKYGMULDALQ--TEG 184
DB 126 TSRTGGILAPVDTMKRABPGKNAIAHYVDKNGMLHALLTPQFFPRELLHDCITRALNEG 185
QY 185 NPATVDEASALELIGKRPKIVGSRPNIKITRPEDLALAQFYM 227
DB 186 -ATTDEASALEYCGFHPQLVGRADNIKVTREPDALAEFYL 227

RESULT 9
US-09-815-242-13867
; Sequence 13867, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 13867
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Salmonella typhi
; NAME/KEY: VARIANT
; LOCATION: (1)....(236)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-815-242-13867

Query Match      41.7%; Score 499.5; DB 9; Length 236;

```



[illegible]

```

RESULT 10
US-09-792-251-8
; Sequence 8, Application US/09792251
; Patent No. US20020160364A1
; GENERAL INFORMATION
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YACM AND YOEJ, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; TITLE OF INVENTION: AND THEIR USE
; FILE REFERENCE: 06286-140001
; CURRENT APPLICATION NUMBER: US/09/792,251
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 232
; TYPE: PR1
; ORGANISM: Bacillus subtilis
; US-09-792-251-8

```

|    | Query Match           | 24.8%  | Score 297     | DB 10 | Length 232                        |
|----|-----------------------|--|---------------|-------|-----------------------------------|
|    | Best Local Similarity | 36.8%  | Pred. 2.5e-22 |       |                                   |
|    | Matches               | 85   | Conservative  | 40    | Mismatches 88; Indels 18; Gaps 9; |
| QY | 10                    | VPPAAGVGMQADRPKQYLPLAGKTVLIENTLTRLLESD-AFQKVAVALSVEDPYMBELS  | 68            |       |                                   |
| Db | 6                     | VIPAAQCGKRMKAKSRNKLPIELKGDPIIHLT-RVFDSSHQCDKILVLINVEGERHFQOL | 64            |       |                                   |
| QY | 69                    | IAXRP---DITTAAGGKERADSVLSATKALBEIAEENMVYVHDAARPCLTSSDHLQTD   | 125           |       |                                   |
| Db | 65                    | LSDYFPQTSILVAGDERQKSHVYKGLAKVK---QEKIVLHDAARPRPK--HEQID      | 116           |       |                                   |
| QY | 126                   | TL--KNDPVG-GIHALSSDHLTKHYDGDITATIDRKAVRALTPQMEKYGMEDALQRT    | 182           |       |                                   |
| Db | 117                   | ELIAFAEQTGAILAVPYKQDITKRQVQDQVSSETIERSISLVAQTPOAFRLSLMKAAHA  | 176           |       |                                   |
| QY | 183                   | EGNAAV--TDESALELL-GHKPKTYEGRPDNKTIRPEELALAQYMQQA             | 231           |       |                                   |
| Db | 177                   | ERKCFILGTDSDLVQWEGGSSVRVVEGSSYTNKILTPDQLTSAELIMSES           | 227           |       |                                   |

```

RESULT 11
US-09-935-943-8
; Sequence 8, Application US/09935943
; Patent No. US20020120963n1
GENERAL INFORMATION:
APPLICANT: Levin, Joshua Z.
APPLICANT: Keglich, Lynette M.
APPLICANT: Budziszewski, Gregory J.
TITLE OF INVENTION: Hetericidic Target Genes and Methods
FILE REFERENCE: Pb/5-31509A
CURRENT APPLICATION NUMBER: US/09/935,943

```

```

; CR REPT FILING DATE: 2001-08-23
; PR CR APPLICATION NUMBER: PCT/EP01/08910
; PR CR FILING DATE: 2001-08-01
; PR CR APPLICATION NUMBER: US 60/222,779
; PR CR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 1: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 302
; TYPE: prt
; ORGANISM: Arabidopsis thaliana
US-05 335-943-8

```

|      |     |                  |              |           |            |    |        |        |      |    |
|------|-----|------------------|--------------|-----------|------------|----|--------|--------|------|----|
| Que  | ✓   | Match            | 23.6%        | Score     | 283.5      | DB | 10     | Length | 302  |    |
| Res  |     | Local Similarity | 34.2%        | Pred. No. | 9e-21      |    |        |        |      |    |
| Mat. | res | 80               | Conservative | 46        | Mismatches | 83 | Indels | 25     | Gaps | 10 |

  

|    |     |                          |                                       |     |
|----|-----|--------------------------|---------------------------------------|-----|
| QY | 10  | VVPAGGCKKMOADPRQYLP      | LAGKTYIEH--TLTLLISDAFOKVAVALISVDPR--  | 63  |
|    |     | :::::                    | :::::                                 |     |
| DB | 82  | TLILGGGCKKKKMSMPQYVPL    | LQPLATLSFFFSMPPE-----VGIIVVDCDPPFR    | 135 |
|    |     | :::::                    | :::::                                 |     |
| QY | 64  | -WPELSIAKHPDITAPGKERADSV | SLAKLEDIASENDWVLWHDAAARCLTGSIDHL      | 122 |
|    |     | :::::                    | :::::                                 |     |
| DB | 136 | IFEEYRESIDVDFALPKERQDSV  | SGLOET-DVNSE--LVCIHSARPLVMTBDE-       | 191 |
|    |     | :::::                    | :::::                                 |     |
| QY | 123 | QIDTLKNDPVGG--TLALSSHD   | TLKAVDGT-ITATIDRKHVRBALTPOMEKYGMLEDA- | 178 |
|    |     | :::::                    | :::::                                 |     |
| DB | 192 | --KYLKKGSAAGAAVLGVPAKAT  | IKENVSJSLVYKTRDKTLMEMQPPQVIRPELLKGF   | 249 |
|    |     | :::::                    | :::::                                 |     |
| QY | 179 | -LQRTKNDPAVITDASALTELGH  | PKPIVBEGRPNDIKITRPELDLALAPYMGQA       | 231 |
|    |     | :::::                    | :::::                                 |     |
| DB | 250 | ELVKSSEG-LEVTIDVSLIVETIK | KEPVYVSQGSYNIKIKTTPDDELLAERILSEDS     | 302 |
|    |     | :::::                    | :::::                                 |     |

```

REST: 12
US-10: 128-713A-6
; Sequence 6, Application US/10128713A
; Publication No. US20030170847A1
; GENETIC INFORMATION:
; APPLICANT: Branucci, Michael G
; TITLE OF INVENTION: Genes Involved in Isoprenoid Compound Production
; F I G REFERENCE: C1-1188
; CURRENT APPLICATION NUMBER: US/10/128,713A
; CURRENT FILING DATE: 2002-04-22
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 6
; LENGTH: 232
; TAG: PRT
; ORGANISM: Rhodococcus erythropolis
US-10: 128-713A-6

```

[illegible]

## RESULT 13

US-10-156-761-11503  
 ; Sequence 11503, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 11503  
 ; LENGTH: 250  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-11503

Query Match 20.5%; Score 245.5; DB 15; Length 250;  
 Best Local Similarity 32.1%; Pred. No. 5.7e-17;

Matches 75; Conservative 34; Mismatches 96; Indels 29; Gaps 6;

QY 9 AVPAAGVGMQADRPKQYLPAGKTVIEHTLTLLSDAFQKAVAI-----SV 59  
 DB 16 AVIPAGRGVRIGPGAPKALRANGTPMLIHAVRAASRAVSLVWVAPEPDTAEVSL 75  
 QY 60 EDPYWEISIAHPDITTPGKERADSVLSALXLEDIASNDWVYHDAARPCLTGSD 119  
 DB 76 LDMH---ALPFTBTVVPGSSROESVTLGLDL--PGCIDVLVHDAARPLVPVDF 128  
 QY 120 IHLQIDTLKNDPVGGLIALSSHTLKV-----DGDITITIDRKHWRALTPQMFYK 172  
 DB 129 VDAIVEAVDGAFAVVPALPLADTVKQVAPAAVPGSEPVATPERALRAVOTQ---- 184  
 QY 173 GMLRDLQRTGEGNA--VTDASALELGHKPKVYEGRPDNIKITRPEDIALAQ 224  
 DB 185 GFDRLTVRAHPTVTDVTDASVWQIGARVVVPGHEAPKTRPDLVLA 238

## RESULT 14

US-10-289-762-621  
 ; Sequence 621, Application US/10289762  
 ; Publication No. US20040006218A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Griffiths, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 ; FILE REFERENCE: 9710-003-999  
 ; CURRENT APPLICATION NUMBER: US/10/289,762  
 ; CURRENT FILING DATE: 2003-03-27  
 ; NUMBER OF SEQ ID NOS: 6849  
 ; SEQ ID NO 621  
 ; LENGTH: 218  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia pneumoniae  
 US-10-289-762-621

Query Match 20.1%; Score 240.5; DB 12; Length 218;  
 Best Local Similarity 32.9%; Pred. No. 1.5e-16;

Matches 74; Conservative 41; Mismatches 92; Indels 19; Gaps 9;

QY 3 PIIQCAVVPAAAGVGMQADRPKQYLPAGKTVIEHTLTLLSDAFQKAVAISSVDP 62

DB 7 FMKSSLLISGGGQTRFGSKIPKQYLPINGTPLVHSL--KLSS--LPQIAEVIWCDP 63

QY 63 YWPEISIAKHEDDITTPGKERADSVLSALXLEDIASNDWVYHDAARPCLTGSDIHL 122

DB 64 SYOR--TQEVYVSPAIR--GERRODSVESGLQOV-----SYPMVLIHGARPPIYDEIR- 115

QY 123 QIDTLKNDPVGGLIALSS--HDTLKHVDGDTITATIDRKHWRALTPQMFYKGLRALAQ 180

DB 116 --DLFTEKXIGARLASPIFYTIKQNP--VNTLDRDNALAIHTQCIKTELREGLA 170

QY 181 -RTGNPAVTDASALELGHKPKVYEGRPDNIKITRPEDIALAQ 224

DB 171 LAKKQULTVVDIRAAHIIIGKPSQLVFNKHPQIKISYEDLTIQA 215

## RESULT 15

US-09-738-626-6422  
 ; Sequence 6422, Application US/09738626  
 ; Publication No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOGUCHI, HIROSHI  
 ; APPLICANT: ANDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIKO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAOKO  
 ; APPLICANT: SENOH, AKIHIRO  
 ; APPLICANT: IKEDA, MASANO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; CURRENT FILING DATE: 2000-12-18  
 ; PRIOR APPLICATION NUMBER: JP 99/377484  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: JP 00/158162  
 ; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: JP 00/280988  
 ; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 7059  
 ; SOFTWARE: PatentIn ver. 3.0  
 ; SEQ ID NO 6422  
 ; LENGTH: 256  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-738-626-6422

Query Match 18.6%; Score 223.5; DB 10; Length 256;  
 Best Local Similarity 30.7%; Pred. No. 1.1e-14;

Matches 71; Conservative 43; Mismatches 102; Indels 15; Gaps 7;

QY 9 AVPAAGVGMQADRPKQYLPAGKTVIEHTLTLLSDAFQKAVAISSVDPYWEI 67  
 DB 10 ALTLAAGRGTRIGPPIKAVTIERLERSIQAMULTSVDHIIIVSPDMETVARDL 69  
 QY 68 ----SIKHDP---IIRPGKERADSVLSALXLE--DISNDWVYHDAARPCLTGSD 119  
 DB 70 LKKGGLNDPVGAVKRLVHGGGEPADSVMAGLOLISDDTDPDAIVLHDSARALTPPGM 129  
 QY 120 IHLQIDTLKNDPVGGLIALSSHTLKV-----DGDITITIDRKHWRALTPQMFYKGLRD 177  
 DB 130 IARVAVKHGATGAVIVLPVPSDTIKRVSDDGVVVTTPRARLRAVOTQGFILSELVA 189  
 QY 178 ALQR--TEGNPAV--TDEASALELGHKPKVYEGRPDNIKITRPEDIALAQ 224  
 DB 190 ANEKFPADPNPGFLTDDASLMEWYGADVVCVQGDPAFAFVTPIDIMLAAQ 240

Search completed: January 29, 2004, 16:21:16  
 Job time : 16.7685 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 : Search time 6.79678 Seconds  
(without alignments)  
3268.453 Million cell updates/sec

Title: US-09-941-947a-10

Perfect score: 1199

Sequence: 1 KMPTIGCAVAVPAAGVCKM.....IKTRPEDLALAQFVMEQQA 231

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 574.5 | 47.9        | 232    | 2     | 2-C-methyl-D-eryth |
| 2          | 555.5 | 46.3        | 234    | 2     | conserved hypotet  |
| 3          | 530.5 | 44.2        | 225    | 2     | hypothetical prote |
| 4          | 524.5 | 43.7        | 236    | 2     | hypothetical prote |
| 5          | 524.5 | 43.7        | 236    | 2     | hypothetical prote |
| 6          | 520.5 | 43.4        | 236    | 2     | 4-diphosphocytidyl |
| 7          | 515.5 | 43.0        | 236    | 2     | 2-C-methyl-D-eryth |
| 8          | 512.5 | 42.7        | 241    | 2     | 4-diphosphocytidyl |
| 9          | 474.5 | 39.6        | 237    | 2     | hypothetical prote |
| 10         | 411.5 | 34.3        | 229    | 2     | conserved hypotet  |
| 11         | 409.5 | 34.2        | 229    | 2     | conserved hypotet  |
| 12         | 378   | 31.5        | 231    | 2     | conserved hypotet  |
| 13         | 353.5 | 29.5        | 231    | 2     | hypothetical prote |
| 14         | 297   | 24.8        | 232    | 2     | conserved hypotet  |
| 15         | 286   | 23.9        | 218    | 2     | conserved hypotet  |
| 16         | 282.5 | 23.6        | 228    | 2     | 4-diphosphocytidyl |
| 17         | 279.5 | 23.3        | 400    | 2     | 4-diphosphocytidyl |
| 18         | 279.5 | 23.3        | 400    | 2     | 4-diphosphocytidyl |
| 19         | 272   | 22.7        | 474    | 2     | CDP-ribitol pyroph |
| 20         | 270   | 22.4        | 474    | 2     | CDP-ribitol pyroph |
| 21         | 268.5 | 22.5        | 232    | 2     | conserved hypotet  |
| 22         | 263   | 21.9        | 232    | 2     | 2-C-methyl-D-eryth |
| 23         | 260.5 | 21.7        | 230    | 2     | hypothetical prote |
| 24         | 254.5 | 21.2        | 231    | 2     | hypothetical prote |
| 25         | 254.5 | 21.2        | 241    | 2     | hypothetical prote |
| 26         | 246   | 20.5        | 236    | 2     | 2-C-methyl-D-eryth |
| 27         | 241.5 | 19.1        | 390    | 2     | 2-C-methyl-D-eryth |
| 28         | 234.5 | 19.6        | 211    | 2     | 2-C-methyl-D-eryth |
| 29         | 234.5 | 19.6        | 211    | 2     | 2-C-methyl-D-eryth |

|    |       |      |     |   |                    |
|----|-------|------|-----|---|--------------------|
| 30 | 233.5 | 19.5 | 229 | 2 | 4-diphosphocytidyl |
| 31 | 227   | 18.9 | 237 | 2 | 2-C-methyl-D-eryth |
| 32 | 227   | 18.9 | 379 | 2 | hypothetical prote |
| 33 | 225   | 18.8 | 382 | 2 | conserved hypotet  |
| 34 | 214.5 | 17.9 | 382 | 2 | conserved hypotet  |
| 35 | 214.5 | 17.9 | 382 | 2 | conserved hypotet  |
| 36 | 211   | 17.6 | 232 | 2 | conserved hypotet  |
| 37 | 209   | 17.4 | 235 | 2 | conserved hypotet  |
| 38 | 209   | 17.4 | 235 | 2 | conserved hypotet  |
| 39 | 201.5 | 16.8 | 218 | 2 | conserved hypotet  |
| 40 | 200.5 | 16.7 | 238 | 2 | hypothetical prote |
| 41 | 190.5 | 15.9 | 399 | 2 | conserved hypotet  |
| 42 | 170.5 | 14.2 | 278 | 2 | 2-C-methyl-D-eryth |
| 43 | 146.5 | 12.2 | 189 | 2 | 2-C-methyl-D-eryth |
| 44 | 141   | 11.8 | 229 | 2 | hypothetical prote |
| 45 | 131.5 | 11.0 | 371 | 2 | hypothetical prote |
|    | 118   | 9.8  | 458 | 2 | hypothetical prote |

## ALIGNMENTS

## RESULT 1

2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60) - Vibrio cholerae

C:Species: Vibrio cholerae

C:Data: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 17-Mar-2003

C:Accession: C82311

R:Heidelberg, J.F.; Bisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chardon, D.; Ermolaeva, M.D.; Vamathevan, U.; Base, S.; Qin, H.; Dragoti, I.; Sellers, P.

1. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: C82311

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-232 <HEI>

A:Cross-references: GB:AE004139; GB:AE003852; NID:9854953; PID:AA93696.1; GSPDB:GN001;

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

A:Genes: VC0528

A:Map position: 1

C:Suprafamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

C:Keywords: nucleotidyltransferase

Query Match 47.9%; Score 574.5; DB 2; Length 232;  
Best Local Similarity 54.4%; Pred. No. 4.5e-41;  
Matches 124; Conservative 28; Mismatches 63; Indels 13; Gaps 5;

|    |     |   |     |
|----|-----|---|-----|
| QY | 9   | AVVPAAGVGRVQADREPKYIPLAGKYIETRLLESDFQKAVASVSDPPWPLS     | 68  |
| DB | 6   | AIVPAAGVGRVQADREPKYIPLAGKYIETRLLESDFQKAVASVSDPPWPLS     | 65  |
| QY | 69  | IAKHPDITAPGKERADSVLSALKALEDIAS--NDVVLVHDARPLGSDHILQIDT  | 126 |
| DB | 66  | LAHPRVIRVNDGGERADSVLS--ALEYVQHLSSEVVLVHDARPCVTADITQLIT  | 122 |
| QY | 127 | LENDPVGILALSSHDTKAYVDG-DTITATIDRKHWVRAALTPQFKYKMD---    | 181 |
| DB | 123 | ALAHFIGALILSPVADTKRQDHLQOIVHTVDRKALMALTPQFRASLEERLPAALQ | 182 |
| QY | 182 | TEGNPAVTVDEASALELGHKPKIVEGRPDNFKITRPEDLALAQFVMEQ        | 229 |
| DB | 183 | ---QVTTIDASAFEMRGERKALVAGRADNFKITOPEDLALAEFLSR          | 227 |

## RESULT 2

conserved hypothetical protein PA3633 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Data: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 17-Mar-2003

C:Accession: F83191

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, J.; Loty, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
 A/Reference number: A82950; MUID:20437337; PMID:10984043  
 A/Accession: F83191  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-234 <STO>  
 A/Cross-references: GB:AE004783; GB:AE004091; NID:G9949786; PIDN:AAG07021.1; GSPDB:GN001  
 A/Experimental source: strain PA01  
 C/Genetics:  
 A/Gene: PA3633  
 C/Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

Query Match 46.3%; Score 555.5; DB 2; Length 234;  
 Best Local Similarity 54.1%; Pred. No. 1.8e-39;  
 Matches 119; Conservative 26; Mismatches 72; Indels 3; Gaps 3;

QY 8 WAVPAAGVGRKQADRPKQYLPPLAKTVEHTLRLLESDFQKAVAVASVEDPYWPEL 67  
 DB 10 WYVPAAGVGRKQADRPKQYLPPLAKTVEHTLRLLESDFQKAVAVASVEDPYWPEL 69  
 QY 68 SIAKHPDITAPGKERADSVLSA-LKALSDIASBDMVTVHDAARPCLTGSDIHQIDTL 126  
 DB 70 DCAASHVGRAGARRASVNLGLRLLEHGAQADMTVTVHDAARPNLTGDDLRLLLE 129  
 QY 127 LKADPVGILALSHDTLKHVDG-TITATIDRKHYWRALTTQMKRYGMLRAL-QRERG 184  
 DB 130 LAEDPVGILAVPARDTLRSDRGRVSTIDRSVWMLAVTQMPFLGALRALDAVA 189  
 QY 185 NPATDEASALELGHKPKYVGRPNIKITRPPDLAQAQ 224  
 DB 190 GVAITDEASALELGHKPKYVGRPNIKITRPPDLAQAQ 229

RESULT 3  
 G64156  
 Hypoetical protein H10672 - Haemophilus influenzae (strain Rd KW20)  
 C/Species: Haemophilus influenzae  
 C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 17-Mar-2003  
 C/Accession: G64156

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kinkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Stitzley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995  
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Smali, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.  
 A/Reference number: A64000; MUID:95350630; PMID:7542800  
 A/Accession: G64156  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-225 <TRIG>  
 A/Cross-references: GB:U32750; GB:I42023; NID:G1573668; PIDN:AAC2332.1; PID:G1573673; T  
 C/Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

Query Match 44.2%; Score 530.5; DB 2; Length 225;  
 Best Local Similarity 52.5%; Pred. No. 2.2e-17;  
 Matches 117; Conservative 28; Mismatches 65; Indels 13; Gaps 6;

QY 9 AVPAAGVGRKQADRPKQYLPPLAKTVEHTLRLLESDFQKAVAVASVEDPYWPEL 68  
 DB 7 AVPAAGVGRKQADRPKQYLPPLAKTVEHTLRLLESDFQKAVAVASVEDPYWPEL 66  
 QY 69 IAKHPDITAPGKERADSVLSA-LKALSDIASBDMVTVHDAARPCLTGSDIHQIDTL 128  
 DB 67 L-DEKIQVEGTTAAESTVINGLAN--LAKNNAVTVHDAARPCLTGSDIHQIDTL 117  
 QY 129 --NDPVGILALSHDTLKHVDG-TITATIDRKHYWRALTTQMKRYGMLRAL-QRERG 184  
 DB 118 AIEDKGAIALPVTDTIRKADQOCIVTEHRSQMLQMTQFPVDTLRDALSTG100 177

QY 185 NPATDEASALELGHKPKYVGRPNIKITRPPDLAQAQ 227  
 DB 178 GANITDEASALELGHKPKYVGRPNIKITRPPDLAQAQ 220

RESULT 4  
 A91079  
 Hypoetical protein Ecs3601 [imported] - *Escherichia coli* (strain O157H7, substrain R1)  
 C/Species: *Escherichia coli*  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-Mar-2003  
 C/Accession: A91079  
 R;Hewari, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasekawa, N.; Yasunaga, T.; Kumara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc  
 A/Reference number: A99629; MUID:21156231; PMID:11258796  
 A/Accession: A91079  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-236 <HAY>  
 A/Cross-references: GB:BA000007; PIDN:BA07024.1; PID:G13363072; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain R1WD 0509952  
 C/Genetics:  
 A/Gene: Ecs3601  
 C/Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

Query Match 43.7%; Score 524.5; DB 2; Length 236;  
 Best Local Similarity 50.2%; Pred. No. 7.6e-37;  
 Matches 112; Conservative 31; Mismatches 71; Indels 9; Gaps 4;

QY 9 AVPAAGVGRKQADRPKQYLPPLAKTVEHTLRLLESDFQKAVAVASVEDPYWPEL 68  
 DB 10 AVPAAGVGRKQADRPKQYLPPLAKTVEHTLRLLESDFQKAVAVASVEDPYWPEL 69  
 QY 69 IAKHPDITAPGKERADSVLSA-LKALSDIASBDMVTVHDAARPCLTGSDIHQIDTL 128  
 DB 70 LANPQITVVDKSGERRASVLAGKA---AGDAQWTVHDAARPCLTGSDIHQIDTL 125  
 QY 129 NDPVGILALSHDTLKHVD--GDTTATIDRKHYWRALTTQMKRYGMLRAL-QRERG 184  
 DB 126 TSPGQILAAVVRDPMKRAEGRKNAIAHTVDRNGMALHTQFPRLHDCITRALNEG 185  
 QY 185 NPATDEASALELGHKPKYVGRPNIKITRPPDLAQAQ 227  
 DB 186 AAITDEASALELGHKPKYVGRPNIKITRPPDLAQAQ 227

RESULT 5  
 B85924

Hypoetical protein ygbp [imported] - *Escherichia coli* (strain O157H7, substrain EDL933)  
 C/Species: *Escherichia coli*  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-Mar-2003  
 C/Accession: B85924  
 R;Perma, N.T.; Plunkett III, G.; Burland, V.; Mu, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 111; L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamou, K.; Apodaca, N. Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: B85924  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-236 <STO>  
 A/Cross-references: GB:AE005174; NID:G12517201; PIDN:AAG57854.1; GSPDB:GN00145; UMGF:240  
 A/Experimental source: strain O157:H7, substrain BDL933  
 C/Genetics:  
 A/Gene: ygbp  
 C/Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

Query Match 43.7%; Score 524.5; DB 2; Length 236;  
 Best Local Similarity 50.2%; Pred. No. 7.6e-37;  
 Matches 112; Conservative 31; Mismatches 71; Indels 9; Gaps 4;

QY 9 AVPAAGVGRKQADRPKQYLPPLAKTVEHTLRLLESDFQKAVAVASVEDPYWPEL 68

```

Db      10 AVPAAGGRRMOTCECPKOYLSIGNOTLHSHVALLAHPRVKRVVLAISGDSRFAPLP 69
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDWLVHDAARPCLTGSDIHLQIDTLK 128
      70 LANHQITVVGGERADSVLAGLKA----AGDAQWLVHDAARPCLTGSDIHLQIDTLK 125
      129 NDPVGGILALSSHTLKVND--GDTTATIDRKHWYRALTPQMKYGMRLDALQR--TEG 184
      126 TSTRGTGIIAAPVDDTMKKAEPGKNAIAHTVDRMGIMHALLTPQFPFRELHDLCTRALNEG 185
      185 NPAYTDEASALELGHKPKIVEGRPNKIKTRPEDIALAQPFM 227
      186 -ATTIDBASALEYCGFHPQVVEGRADNIKVTRPEDIALAEFTL 227

RESULT 6
665055
4-diphosphocytidyl-2C-methyl-D-erythritol synthase (EC 2.7.7.-) (similarity) - Escherich
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2003
C:Accession: G65055
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
-A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65055
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-236 <3LAT>
A:Cross-references: GB:AE000358; GB:U00096; NID:92367156; PIDN:AACT5789.1; PID:gl789104;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase
C:Keywords: nucleotidyltransferase

Query Match      43.4%; Score 520.5; DB 2; Length 236;
Best Local Similarity 49.8%; Pred. No. 1.7e-36;
Matches 111; Conservative 31; Mismatches 72; Indels 9; Gaps 4;

Db      9 AVPAAGGCKMQADRPKOYLPLAKTVIEHTLRLLESADAFQKAVASVEDPYWELS 68
      10 AVPAAGGRRMOTCECPKOYLSIGNOTLHSHVALLAHPRVKRVVLAISGDSRFAPLP 69
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDWLVHDAARPCLTGSDIHLQIDTLK 128
      70 LANHQITVVGGERADSVLAGLKA----AGDAQWLVHDAARPCLTGSDIHLQIDTLK 125
      129 NDPVGGILALSSHTLKVND--GDTTATIDRKHWYRALTPQMKYGMRLDALQR--TEG 184
      126 TSTRGTGIIAAPVDDTMKKAEPGKNAIAHTVDRMGIMHALLTPQFPFRELHDLCTRALNEG 185
      185 NPAYTDEASALELGHKPKIVEGRPNKIKTRPEDIALAQPFM 227
      186 -ATTIDBASALEYCGFHPQVVEGRADNIKVTRPEDIALAEFTL 227

RESULT 7
AE0856
2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (imported) - Salmonella enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: This species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 17-Mar-2003
C:Accession: AE0856
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Mole, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Farry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0856
A>Status: preliminary

```

```

A:Molecule type: DNA
A:Residues: 1-236 <PAR>
A:Cross-references: GB:AL533382; PIDN:CAD06036.1; PID:gl6504003; GSPDB:GN00176
C:Genetics:
A:Gene: ysbP
C:Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase

Query Match      43.0%; Score 515.5; DB 2; Length 236;
Best Local Similarity 49.8%; Pred. No. 4.4e-36;
Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;

Db      9 AVPAAGGCKMQADRPKOYLPLAKTVIEHTLRLLESADAFQKAVASVEDPYWELS 68
      10 AVPAAGGRRMOTCECPKOYLSIGNOTLHSHVALLAHPRVKRVVLAISGDSRFAPLP 69
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDWLVHDAARPCLTGSDIHLQIDTLK 128
      70 LANHQITVVGGERADSVLAGLQAV---AKAQWLVHDAARPCLTGSDIHLQIDTLK 125
      129 NDPVGGILALSSHTLKVND--GDTTATIDRKHWYRALTPQMKYGMRLDALQR--TEG 184
      126 NSRVGGILASPVDTMKKAEPGKNAIAHTVDRADWALTPQFPFRELHDLCTRALNEG 185
      185 NPAYTDEASALELGHKPKIVEGRPNKIKTRPEDIALAQPFM 227
      186 -ATTIDBASALEYCGFHPALVEGRADNIKVTRPEDIALAEFTL 227

RESULT 8
AC0406
4-diphosphocytidyl-2C-methyl-D-erythritol synthase (EC 2.7.7.-) (imported) - Yersinia pes
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-Mar-2003
C:Accession: AC0406
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
demo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0406
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <KUR>
A:Cross-references: GB:AL590842; PIDN:CA092591.1; PID:gl5981288; GSPDB:GN00175
C:Genetics:
A:Gene: ispD
C:Superfamily: 4-diphosphocytidyl-2-methyl-D-erythritol synthase
C:Keywords: nucleotidyltransferase

Query Match      42.7%; Score 512.5; DB 2; Length 241;
Best Local Similarity 49.1%; Pred. No. 8e-36;
Matches 111; Conservative 35; Mismatches 71; Indels 9; Gaps 4;

Db      9 AVPAAGGCKMQADRPKOYLPLAKTVIEHTLRLLESADAFQKAVASVEDPYWELS 68
      13 AVPAAGGRRMOTCECPKOYLTGKTIHAFSLHBRIRGVIVHPDQFSRLS 72
      69 IAKHPDITAPGKERADSVLSALKALEDIASENDWLVHDAARPCLTGSDIHLQIDTLK 128
      73 VADQPRISTYGGDQGRANSMAGIQ----LAQAEMWLVHDAARPCLTGSDIHLQIDTLK 128
      129 NDPVGGILALSSHTLKVND--GDTTATIDRKHWYRALTPQMKYGMRLDALQR--TEG 184
      129 CSQVGGIIAPVDDTMKKAEPGKNAIAHTVDRADWALTPQFPFRELHDLCTRALNEG 188
      185 NPAYTDEASALELGHKPKIVEGRPNKIKTRPEDIALAQPFM 230
      189 -VAVTDEASALEHCGFHPILVTRGSNDNIKVTRPEDIALAEFTLQR 233

RESULT 9
F84978

```



Db 120 ETAAKTH-AAVLAVPVKDTIKRVGEAVLETMPREELNAVQTPQAFDLAL

Qy 185 NPAV-TDEASALELGHKPKIVBGRDNIKITRPEDIALAQFYMEQ 23C

Dd 177 ERKGLGTDDA\$LVQMEGGSVRVVEGSYTNIKLITPPDDL\$A\$AIMESES 227

```

RESULT 15
D70414
Conserved hypothetical protein aq_1323 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #next_change 17-Mar-2003
C:Accession: D70414
C:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overman, L.B.

```



V.

Mature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; KUID:98198566; PMID:9537320

A:Accession: D70414

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-213 &lt;AOF&gt;

A:Cross-references: GB:AE000734; NID:G2983733; PIDN:AA007307.1; PID:G2983746; GB:AE00066

A:Experimental source: strain VFS

A:Genetics:

A:Gene: aq\_1323

C:Superfamily: 4-diphosphocytidylyl-2-methyl-D-erythritol synthase

Query Match

23.9%; Score 286; DB 2; Length 213;

Best Local Similarity 35.4%; Pred. No. 8, 7e-17;

Matches 79; Conservative 40; Mismatches 82; Indels 22; Gaps 8;

```

QY      9 AVVPAAGVGNQADRPQYLPAGKTVIENTLRLESDFQKVAVAISVEDPYWPELS 68
DB      4 AIIAAGSGSRI--GFRKQFATLCGKPLFMHSLERKL--DIFEEVILVL-----PEDF 52
QY      69 IAK--HPDIITAPGKXERADSVLSALALMEDIASENDVYVHDARPCLTGSDIHLQID 125
DB      53 LDKVKVHPKVKVAGGPERQDSVFNAL-----LQATGDIVIHDSARPLAT--KMELEVA 106
QY      126 TKNDPVGGLALSSHDTLKHVDGDTITATIDRKHWRALTPQWFKYGLRDALQRTEN 185
DB      107 QL-GDYHGKVVASPARDTLKEVEGKVIKTLNRSIIWHAQTPOAFRRDILLECHWRAYAE 165
QY      186 EAV-TDEASALDELHKPKIVGGRPDNLIKITRPEDLALAQFTM 227
DB      166 GFVGTDDASLIERGYSGVVEGSYWNVKITYPEDLIEWYKIM 208

```

Search completed: January 29, 2004, 15:55:30

Job time : 6.79678 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 4.34894 Seconds  
(without alignments)  
2497.314 Million cell updates/sec

Title: US-09-941-947A-10

Sequence: 1 MNPTQCNAVVPAGVGRM.....IKITRPDIALAGFMGQA 231

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                     |
|------------|-------|-------------|--------|-------|---------------------------------|
| 1          | 605   | 50.5        | 237    | 1     | ISPD_VIBVU Q8460 vibrio vuln    |
| 2          | 594.5 | 49.6        | 234    | 1     | ISPD_VIBRA Q871q2 vibrio para   |
| 3          | 574.5 | 47.9        | 232    | 1     | ISPD_VIBCH Q9KUJ2 vibrio chol   |
| 4          | 555.5 | 46.3        | 234    | 1     | ISPD_PSEAE P57707 pseudomonas   |
| 5          | 530.5 | 44.2        | 225    | 1     | ISPD_HAIAI O08029 haemophilus   |
| 6          | 524.5 | 43.7        | 235    | 1     | ISPD_ECO57 O847Y4 escherichia   |
| 7          | 520.5 | 43.4        | 235    | 1     | ISPD_ECO16 O847E5 escherichia   |
| 8          | 520.5 | 43.4        | 235    | 1     | ISPD_ECO11 Q84893 escherichia   |
| 9          | 515.5 | 43.0        | 236    | 1     | ISPD_SALT1 Q84771 salmonella    |
| 10         | 513.5 | 42.8        | 236    | 1     | ISPD_SALT1 Q84771 salmonella    |
| 11         | 512.5 | 42.7        | 241    | 1     | ISPD_YERPE Q82b6 yersinia pe    |
| 12         | 505   | 42.1        | 238    | 1     | ISPD_PASNU P57953 pasteurella   |
| 13         | 474.5 | 39.6        | 237    | 1     | ISPD_BUCAI P57495 buchnera ap   |
| 14         | 462.5 | 38.6        | 229    | 1     | ISPD_WIGBR Q84223 wiglesworth   |
| 15         | 457.5 | 38.2        | 236    | 1     | ISPD_BUCAP Q849d6 buchnera ap   |
| 16         | 425   | 35.4        | 265    | 1     | ISPD_XANCP Q85921 xanthomonas   |
| 17         | 422   | 35.2        | 266    | 1     | ISPD_XANCP Q85921 xanthomonas   |
| 18         | 411.5 | 34.3        | 229    | 1     | ISPD_NEIMA Q91tm3 neisseria m   |
| 19         | 409.5 | 34.2        | 229    | 1     | ISPD_NEIMA Q91tm3 neisseria m   |
| 20         | 398   | 33.2        | 253    | 1     | ISPD_RALSO Q8xyw3 ralsostonia s |
| 21         | 378   | 31.5        | 231    | 1     | ISPD_YEYFA Q9d6f8 xyloella fas  |
| 22         | 353.5 | 28.5        | 228    | 1     | ISPD_BACHD Q9d6f8 xyloella fas  |
| 23         | 301   | 25.1        | 232    | 1     | ISPD_BACSU Q8755 bacillus su    |
| 24         | 286   | 23.9        | 213    | 1     | ISPD_AQUAE Q87343 aquifex aeo   |
| 25         | 282.5 | 23.6        | 228    | 1     | ISPD_ANASP Q871x9 anabaena ap   |
| 26         | 279.5 | 23.3        | 400    | 1     | ISPD_AGRIS Q8uf4 a ispd/ispf    |
| 27         | 270   | 22.5        | 232    | 1     | ISPD_LISMO Q8yab5 listeria mo   |
| 28         | 268.5 | 22.4        | 222    | 1     | ISPD_THEMA Q8yab5 listeria mo   |
| 29         | 263   | 21.9        | 231    | 1     | ISPD_TUSNA Q8yab5 listeria mo   |
| 30         | 263   | 21.9        | 232    | 1     | ISPD_LISIN Q8yab5 listeria mo   |
| 31         | 260.5 | 21.7        | 230    | 1     | ISPD_SYNY3 P74323 cytochrome c  |
| 32         | 256   | 21.4        | 225    | 1     | ISPD_CLOPE Q8xh3 clostridium    |
| 33         | 254.5 | 21.2        | 231    | 1     | ISPD_MYCTU P86864 mycobacteri   |

|    |       |      |     |   |                                |
|----|-------|------|-----|---|--------------------------------|
| 34 | 254.5 | 21.2 | 241 | 1 | ISPD_MYCTE Q9cwe6 mycobacteri  |
| 35 | 247   | 20.6 | 227 | 1 | ISPD_THETN Q8786 thermoanaer   |
| 36 | 247   | 20.6 | 407 | 1 | ISPD_RHTLO Q98mx9 r ispd/ispf  |
| 37 | 243.5 | 20.3 | 270 | 1 | ISPD_STRICO Q910q8 streptomyce |
| 38 | 241.5 | 20.1 | 390 | 1 | ISPD_BRUME Q8yhd8 b ispd/ispf  |
| 39 | 234.5 | 19.6 | 211 | 1 | ISPD_CHLPM Q927x5 chlamydia p  |
| 40 | 233.5 | 19.5 | 229 | 1 | ISPD_CLOAB Q976c9 clostridium  |
| 41 | 227   | 18.9 | 379 | 1 | ISPD_RHOCA Q88113 r ispd/ispf  |
| 42 | 225   | 18.8 | 382 | 1 | ISPD_CAOCR Q8a75 c ispd/ispf   |
| 43 | 225   | 18.8 | 434 | 1 | ISPD_RHIME Q92q90 r ispd/ispf  |
| 44 | 223.5 | 18.6 | 256 | 1 | ISPD_CORGL Q8nmb8 corynebacte  |
| 45 | 211   | 17.6 | 232 | 1 | ISPD_DRIRA Q9r30 deinococcus   |

## ALIGNMENTS

| RESULT 1 | ID   | ISPD_VIBVU                          | STANDARD | PRT | 237 AA |
|----------|--|-------------------------------------|----------|-----|--------|
| AC       | Q8460  |                                     |          |     |        |
| DT       | 15-SEP-2003  | (Rel. 42, Last sequence update)     |          |     |        |
| DT       | 15-SEP-2003  | (Rel. 42, Last sequence update)     |          |     |        |
| DT       | 15-SEP-2003  | (Rel. 42, Last sequence update)     |          |     |        |
| DE       | 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)   |                                     |          |     |        |
| DE       | 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (MSP  |                                     |          |     |        |
| DE       | cytidyltransferase) (MCT)  |                                     |          |     |        |
| GN       | ISPD OR VY11582.   |                                     |          |     |        |
| OC       | Vibrio vulnificus.   |                                     |          |     |        |
| OC       | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  |                                     |          |     |        |
| OX       | Vibrionaceae; Vibrio.  |                                     |          |     |        |
| OX       | NCBI_TaxID=672;  |                                     |          |     |        |
| RM       | [1]  |                                     |          |     |        |
| RP       | SEQUENCE FROM N.A.   |                                     |          |     |        |
| RC       | STRAIN=CMCP6;  |                                     |          |     |        |
| RA       | Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.   |                                     |          |     |        |
| RA       | Choi H.E.;   |                                     |          |     |        |
| RT       | "Complete genome sequence of Vibrio vulnificus CMCP6."   |                                     |          |     |        |
| RL       | Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.  |                                     |          |     |        |
| CC       | -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-  |                                     |          |     |        |
| CC       | methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-   |                                     |          |     |        |
| CC       | phosphate (by similarity).   |                                     |          |     |        |
| CC       | -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  |                                     |          |     |        |
| CC       | diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.   |                                     |          |     |        |
| CC       | -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.   |                                     |          |     |        |
| CC       | -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  |                                     |          |     |        |
| CC       | This SWISS-PROT entry is copyright. It is produced through a collaboration   |                                     |          |     |        |
| CC       | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |                                     |          |     |        |
| CC       | the European Bioinformatics Institute. There are no restrictions on its  |                                     |          |     |        |
| CC       | use by non-profit institutions as long as its content is in no way   |                                     |          |     |        |
| CC       | modified and this statement is not removed. Usage by and for commercial  |                                     |          |     |        |
| CC       | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |                                     |          |     |        |
| CC       | or send an email to <a href="mailto:license@sb-sib.ch">license@sb-sib.ch</a> ).  |                                     |          |     |        |
| CC       | -----  |                                     |          |     |        |
| CC       | EMBL; AB016802; AA010005.1; -  |                                     |          |     |        |
| DR       | HAVAP; MF_00108; -; 1.   |                                     |          |     |        |
| DR       | InterPro; IPR001228; ISPD_synthase.  |                                     |          |     |        |
| DR       | Pfam; PF01128; ISPD; 1.  |                                     |          |     |        |
| DR       | TIGRFAMs; TIGR00453; ispd; 1.  |                                     |          |     |        |
| DR       | PROSITE; PS01295; ISPD; 1.   |                                     |          |     |        |
| KW       | Transferase; Nucleotidyltransferase; Isoprene biosynthesis;  |                                     |          |     |        |
| KW       | Complete proteome.   |                                     |          |     |        |
| SO       | SEQUENCE 237 AA; 26005 MW; 35D0952D12847BCD CRC64;   |                                     |          |     |        |
| Qy       | Query Match  | 50.5%; Score 605; DB 1; Length 237; |          |     |        |
| Db       | Best Local Similarity  | 55.4%; Pred. No. 9.6e-46;           |          |     |        |
|          | Matches 124; Conservative 31; Mismatches 67; Indels 2; Gaps 2;   |                                     |          |     |        |
|          | 9 AAVPAAGVGRMADREKQVLPPLAKGVIVETHTLRLLESAPKQKVAIVSVEQVPELS 68  |                                     |          |     |        |
|          | 10 AAVPAAGVGRMADREKQVLPPLAKGVIVETHTLRLLESAPKQKVAIVSVEQVPELS 69   |                                     |          |     |        |

QY 69 IAKHDIITAGGKXRADSVLSALKALBEDIASENDWVWHDARPCLTGSDIHQIDTLK 128  
 DB 70 IAKHDIIVAVAGKXRADSVLSALKALBEDIASENDWVWHDARPCLTGSDIHQIDTLK 129  
 QY 129 NDPVGGIILASHTLKHVGD--TITATIDRKHWRAALTPQMEKXGMRLDAL--QRTGPNP 186  
 DB 130 SHETGGIILATPVDITMKRANAGQXIDHTYDRNALHMLATPQMEKXGAVLTDLSDALAGV 189  
 QY 187 ATDEASALBELLGKPKIVGSRPNIKITREPDLALAQFYMEQ 230  
 DB 190 ATDEASALEMRGELPALVQGSNNIKYQPEDLALAEFYLSRE 233

## RESULT 2

ISPD\_VIBPA STANDARD; PRT; 234 AA.  
 AC Q87I02;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (BC 2.7.7.60)  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP  
 DE cytidyltransferase) (MCT).  
 GN ISPD OR VP2559.  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 NCBI\_TaxID=670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIIMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Naito M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shingawa H., Hatgori M., Iida T.,  
 RA "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 RT distance from that of V. cholerae.";  
 RL Lancet 361:743-749 (2003).  
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidylyl-2C-  
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-  
 CC phosphate (By similarity).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AP005081; BAC60822.1; -  
 DR HAMAP: MF\_00108; 1  
 DR PROSITE: PS01293; ISPD, 1.  
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;  
 DR Complete proteome.  
 SK SEQUENCE 234 AA; 25470 MW; E92046C6D8C0682 CRC64;

Query Match 49.6%; Score 594.5; DB 1; Length 234;  
 Best Local Similarity 55.9%; Pred. No. 7, 9e-45;  
 Matches 127; Conservative 30; Mismatches 61; Indels 9; Gaps 4;

QY 9 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLRLIESDAFQKVAVAISVEDPYPELS 68  
 DB 10 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLRLIESDAFQKVAVAISVEDPYPELS 69  
 QY 69 IAKHDIITAGGKXRADSVLSALKALBEDIASENDWVWHDARPCLTGSDIHQIDTLK 128  
 DB 70 IAKHDIIVAVAGKXRADSVLSALKALBEDIASENDWVWHDARPCLTGSDIHQIDTLK 128

QY 129 NDPVGGIILASHTLKHVGD--TITATIDRKHWRAALTPQMEKXGM-----IIPALORTE 183  
 DB 129 AHPFGGIIASVTRITMKRANAGQXIDHTYDRNALHMLATPQMEKXGAVLTDLSDALAGV 186  
 QY 184 GNPATVDEASALBELLGKPKIVGSRPNIKITREPDLALAQFYMEQ 230  
 DB 187 -GVATIDBASALEMRGELPALVQGSNNIKYQPEDLALAEFYLSRE 232

## RESULT 3

ISPD\_VIBCH STANDARD; PRT; 232 AA.  
 AC Q9KUT2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (BC 2.7.7.60)  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP  
 DE cytidyltransferase) (MCT).  
 GN ISPD OR VC0528.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI For M1961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickley E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Rischard D.,  
 RA Ermolenko M.D., Vamathevan J., Bass S., Qin H., Diracoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Newman M.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.,  
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483 (2000).  
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-  
 CC METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-  
 CC PHOSPHATE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE004139; AAF93696.1; -  
 DR PIR: C82311; C82311.  
 DR HSSP: Q46893; 1152.  
 DR TIGR: VC0528; -  
 DR HAMAP: MF\_00108; 1  
 DR InterPro: IPR001228; ISPD\_synthase.  
 DR Pfam: PF01128; ISPD, 1.  
 DR PIRSF: PIRSF006765; DPCMB\_synth; 1.  
 DR TIGRFAMs: TIGRF00453; ISPD, 1.  
 DR PROSITE: PS01293; ISPD, 1.  
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;  
 DR Complete proteome.  
 SK SEQUENCE 232 AA; 25901 MW; 44AA02420C724354 CRC64;

Query Match 47.8%; Score 574.5; DB 1; Length 232;  
 Best Local Similarity 54.4%; Pred. No. 4, 4e-43;  
 Matches 124; Conservative 28; Mismatches 63; Indels 13; Gaps 5;

QY 9 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLRLIESDAFQKVAVAISVEDPYPELS 68

DB 6 AIVPAAGVGRNQADRPQYLTLDKTYLLEHTVEHLLEHPLLEHVAVASADDFPALP 65  
 QY 69 IAKEDPITTPGKERRADSVLSALALEDIAS--NDVYLVDAAAPCTGSDIHLQIDT 126  
 DB 66 LAHFRVIRIVDGGKERRASVLS---ALRYVCOHRLESEWLVVDAAAPCTVHADITQIIT 122  
 QY 127 LKNDVPGGIALSSDPLKRVDCD-TTATATDRKVRALTPQMKYGMALD---ALQR 181  
 DB 123 ALAHFICGIALSPVADPMKRGDHLQOIVHTVDRKLMHMLTPQMRAGSLERLFPALQ 182  
 QY 182 TEGNDVATDEASALELGHKPKIVEGRPNIKITRPEDLALAQFYMEQ 229  
 DB 183 ---QVITDEASAFEMRGKFPALVAGRADNMLKITRPEDLALAEFLSR 227

## RESULT 4

ISPD\_PSEAE STANDARD; PRT; 234 AA.  
 AC P57707; Q9HXZ7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyllyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP  
 CYTIDYLYLTRANSFERASE) (MCT).  
 GN ISPD OR PA3613.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 NCBI\_TaxID=287;  
 RN (1)

SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Phan X.-Q.T., Eyrin A.L., Mizoguchi S.D., Warren P.,  
 Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 Garber R.L., Goltier S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,  
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,  
 Smith K.A., Spencer D.H., Wong G.R.-S., Wu Z., Paulsen I.T.,  
 Reiter J., Sater M.H., Hancock R.R.W., Lory S., Olson M.V.;  
 RA Complete genome sequence of Pseudomonas aeruginosa PAOI, an  
 opportunistic pathogen.  
 RL Nature 406:959-964(2000).  
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-  
 METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-  
 PHOSPHATE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 -1- diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonaevalonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB004783; AAC07021.1; -  
 DR PIR; F83191; F83191.  
 DR HSSP; Q46893; 1152.  
 DR HAMAP; MF\_00108; -; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR PIRSF; PIRSF006765; DPCME\_synth; 1.  
 DR TIGRFAMs; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;  
 KM Complete proteome.  
 SQ SEQUENCE 234 AA; 25613 MF; 2CBA037CD85B42C9 CIRC64;

Query Match 46.3%; Score 555.5; DB 1; Length 234;  
 Best Local Similarity 54.1%; Pred. No. 2e-41;  
 Matches 119; Conservative 26; Mismatches 72; Indels 3; Gaps 3;  
 QY 8 WAVPAAGVGRNQADRPQYLTPLAGTVEHTLTLLESDAPKVAIVASVEDPYMPL 67  
 DB 10 WTVPAAGVGRNQADRPQYLTDLAGTVEHTLTLLESDAPKVAIVASVEDPYMPL 69  
 QY 68 SIAGHPDITTPGKERRADSVLS--LKALEDIASNDVYLVDAAAPCTGSDIHLQIDT 126  
 DB 70 DCAASHVQRAAGAEAGSVLSNGLRLLEGAQADWVYVHDAAAPNTLRGDLRLLE 129  
 QY 127 LKNDVPGGIALSSDPLKRVDCD-TTATATDRKVRALTPQMKYGMALD---QRTG 184  
 DB 130 LAEDPVGGLAVPARDLTKSDRCGRVSEITIDRSVVLATTPQMRGALHRLADALVA 189  
 QY 185 NPATDEASALELGHKPKIVEGRPNIKITRPEDLALAQ 224  
 DB 190 GVAITDEASAFEMRGKFPALVAGRADNMLKITRPEDLALAEFLSR 229

## RESULT 5

ISPD\_HAIRIN STANDARD; PRT; 225 AA.  
 AC Q05029;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyllyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP  
 CYTIDYLYLTRANSFERASE) (MCT).  
 GN ISPD OR H10672.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 NCBI\_TaxID=727;  
 RN (1)

SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 Karlyavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 McEwen K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 Uccerack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghagen N.S.M.,  
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 Venter J.C.;  
 RA Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT R1.";  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-  
 D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-PHOSPHATE (BY  
 SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 -1- diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonaevalonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U32750; AAC22332.1; -  
 DR PIR; G64156; G64156.  
 DR HSSP; Q46893; 1152.  
 DR TIGR; H10672; -  
 DR HAMAP; MF\_00108; -; 1.

DR InterPro: IPR001228; ISPD\_synthase.  
 DR Pfam: PF01128; ISPD; 1.  
 DR PIRSF: PIRSF006765; DPCME\_synth; 1.  
 DR TIGRFAMs: TIGR00453; ISPD; 1.  
 DR PROSITE: PS01295; ISPD; 1.  
 DR Trasnase; Nucleotidyltransferase; Isoprene biosynthesis;  
 DR Complete proteome.  
 SQ SEQUENCE 225 AA; 24548 MW; 2155A297B44CD8C9 CRC64;  
 Query Match 44.2%; Score 530.5; DB 1; Length 225;  
 Best Local Similarity 52.5%; Pred. No. 2,9e-39;  
 Matches 117; Conservative 28; Mismatches 65; Indels 13; Gaps 6;  
 QY 9 AVVPAAGVGRMQRADPKQYLPVLAAGTVVHTLTLTLLSDAPQKVAVAISVEDPYWELLS 68  
 DB 7 AVVPAAGVGRMQRADPKQYLPVLAAGTVVHTLTLTLLSDAPQKVAVAISVEDPYWELLS 66  
 QY 69 IAKHPDITAPGGERADSVLSALKALIEDIASENDVAVDAARPCLTGSDIHLQIDTLK 128  
 DB 67 L--DPEKIQVLEGQTTAEVSLNGLNA---IAEKNAVAVHDAARPCLTGSDIHLQIDTLK 117  
 QY 129 --NDPVGILALSHDTLKHVDG-DITATIDRKHWRAALPQWFKYGLMDALQOR--TEG 184  
 DB 118 AIEDKQAGLALPVDTITKADNQCIVKTESGQLWQMTQPPVDILADALSTGIQ 177  
 QY 185 NPATVDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYM 227  
 DB 178 GANITDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYL 220  
 RESULT 6  
 ISPD\_ECO57 STANDARD; PRT; 235 AA.  
 AC 08x774;  
 DT 28-PEB-2003 (Rel. 41, Last sequence update)  
 DT 28-PEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyllyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP  
 DE Cytidyllyltransferase) (MCT) (CDP-ME synthetase).  
 GN ISPD OR Z4055 OR ECS3601.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=63334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postell G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Ditalanta E.T., Potamocis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RT Nature 403:529-533 (2001).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
 RA Kohara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22 (2001).  
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-methyl-  
 CC D-erythritol from CDP + 2C-methyl-D-erythritol 4-phosphate =  
 CC -1- CATALYTIC ACTIVITY: CDP + 2-C-methyl-D-erythritol 4-phosphate =  
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- COFACTOR: Magnesium, manganese or cobalt (By similarity).

CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 DR EXML: AE005502; AAG57854.1;  
 DR EXML: AP002562; BA837024.1;  
 DR PIR: A91079; A91079.  
 DR PIR: B85924; B85924.  
 DR HAMAP: MP\_00108; 1.  
 DR InterPro: IPR001228; ISPD\_synthase.  
 DR Pfam: PF01128; ISPD; 1.  
 DR PIRSF: PIRSF006765; DPCME\_synth; 1.  
 DR TIGRFAMs: TIGR00453; ISPD; 1.  
 DR PROSITE: PS01295; ISPD; 1.  
 DR Trasnase; Nucleotidyltransferase; Isoprene biosynthesis; Magnesium;  
 DR Complete proteome.  
 DR Manganese; Cobalt; Complete proteome.  
 FT INIT MET 0  
 SQ SEQUENCE 235 AA; 25602 MW; 3909048B4842ALDC CRC64;  
 Query Match 43.7%; Score 524.5; DB 1; Length 235;  
 Best Local Similarity 50.2%; Pred. No. 1e-38;  
 Matches 112; Conservative 31; Mismatches 71; Indels 9; Gaps 4;  
 QY 9 AVVPAAGVGRMQRADPKQYLPVLAAGTVVHTLTLTLLSDAPQKVAVAISVEDPYWELLS 68  
 DB 9 AVVPAAGVGRMQRADPKQYLPVLAAGTVVHTLTLTLLSDAPQKVAVAISVEDPYWELLS 68  
 QY 69 IAKHPDITAPGGERADSVLSALKALIEDIASENDVAVDAARPCLTGSDIHLQIDTLK 128  
 DB 69 IAKHPDITAPGGERADSVLSALKALIEDIASENDVAVDAARPCLTGSDIHLQIDTLK 124  
 DB 69 IAKHPDITAPGGERADSVLSALKALIEDIASENDVAVDAARPCLTGSDIHLQIDTLK 124  
 QY 129 NDPVGILALSHDTLKHVD-GDTTATIDRKHWRAALPQWFKYGLMDALQOR--TEG 184  
 DB 125 TSRTGILALPVDTITKADNQCIVKTESGQLWQMTQPPVDILADALSTGIQ 184  
 QY 185 NPATVDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYM 227  
 DB 185 NPATVDEASALBELGKPKIVBGRPNIKITRPEDLALAQFYL 226  
 RESULT 7  
 ISPD\_ECO57 STANDARD; PRT; 235 AA.  
 AC 08x774;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyllyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP  
 DE Cytidyllyltransferase) (MCT).  
 GN ISPD OR C3314.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22386234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,  
 RA Mayhew G.F., Rose D.J., Zhou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

CC - FUNCTION: Catalyzes the formation of 4-diphosphocytidylyl-2C-  
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-  
 CC phosphate (By similarity).  
 CC - CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 CC - COFACTOR: Magnesium, manganese or cobalt (By similarity).  
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC - SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB016765; AAN81763.1; -.  
 DR HAMAP; MF\_00108; -; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR TIGRFAMs; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis; Magnesium;  
 KW Manganese; Cobalt; Complete proteome.  
 FT INIT MET 0 By SIMILARITY.  
 SQ SEQUENCE 235 AA; 25648 MW; A4DD31D950D6CE6F CRC64;  
 Query Match 43.4%; Score 520.5; DB 1; Length 235;  
 Best Local Similarity 49.8%; Pred. No. 2,3e-38;  
 Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;  
 QY 9 AAVPAAGCERNQADRPKQYLPAGKTVIEFTLRLESDAFOKAAVAISVDPYWPBLS 68  
 DB 9 AAVPAAGCERNQADRPKQYLPAGKTVIEFTLRLESDAFOKAAVAISVDPYWPBLS 68  
 QY 69 IAKHPDIIITAPGKREKADSVLSAKALBEDIASENMVWVHDARCLTSGDHLIDILK 128  
 DB 69 IAKHPDIIITAPGKREKADSVLSAKALBEDIASENMVWVHDARCLTSGDHLIDILK 128  
 QY 129 NDVPGGIALSSHDITLKHVD--GDITITATIDRKHWRAITPQFKYGMKDALQR--TEG 184  
 DB 129 NDVPGGIALSSHDITLKHVD--GDITITATIDRKHWRAITPQFKYGMKDALQR--TEG 184  
 QY 125 TSRTGGLAAPVDITKRAIPGKMAIAHVVDNGMLHALTPQFFREILHDCITALNMG 184  
 DB 125 TSRTGGLAAPVDITKRAIPGKMAIAHVVDNGMLHALTPQFFREILHDCITALNMG 184  
 QY 185 NPAYVDEASALELGGKKEIVGRPNITCTRPDLAAQFTM 227  
 DB 185 NPAYVDEASALELGGKKEIVGRPNITCTRPDLAAQFTM 227  
 QY 185 -ATTIDBASALEYCGHPQLVEGRADNIKVTREPDLALAEFTL 226  
 DB 185 -ATTIDBASALEYCGHPQLVEGRADNIKVTREPDLALAEFTL 226  
 RESULT 8  
 ID ISPD\_ECOLI STANDARD; PRT; 235 AA.  
 AC Q46893;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DB 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MBP  
 GN cytidylyltransferase) (MCT) (CDP-ME synthetase).  
 OS ISPD OR B2747.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RC STRAIN=K12 / DHS-alpha;  
 RX MEDLINE=99449755; PubMed=10518523;  
 RA Rohdich F., Wungsintaweekul J., Fellemeier M., Sagner S., Herz S.,  
 RA Kls K., Bisenrich M., Bacher A., Zenk M.H.;  
 RT "Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids: Ygbp  
 RT protein of Escherichia coli catalyzes the formation of 4-  
 RT diphosphocytidylyl-2-C-methylerythritol.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:11758-11763(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RA Kuzuyama T., Takagi M., Kaneda K., Dairi T., Seto H.;  
 RT "Formation of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol from  
 RT 2-C-methyl-D-erythritol 4-phosphate by 2-C-methyl-D-erythritol  
 RT 4-phosphate cytidylyltransferase, a new enzyme in the nonmevalonate  
 RT pathway.";  
 RL Tetrahedron Lett. 41:703-706(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perma N.T., Burland V.,  
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mathew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Kau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.50 ANGSTROMS).  
 RX MEDLINE=21320925; PubMed=11427897;  
 RA Richard S.B., Bowman M.R., Kwiatkowski W., Kang I., Chow C.,  
 RA Hillo A.W., Cane D.B., Noel J.P.;  
 RT "Structure of 4-diphosphocytidylyl-2-C-methylerythritol synthetase  
 RT involved in mevalonate-independent isoprenoid biosynthesis.";  
 RL Nat. Struct. Biol. 8:641-648(2001).  
 CC - FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTDYLYL-2C-METHYL-  
 CC D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-PHOSPHATE.  
 CC - CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 CC - COFACTOR: MAGNESIUM, MANGANESE OR COBALT.  
 CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC - SUBUNIT: Homodimer.  
 CC - MISCELLANEOUS: OPTIMAL PH IS 8.3.  
 CC - SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF230736; AAF43207.1; -.  
 DR EMBL; AB037143; BAA90761.1; -.  
 DR EMBL; U29579; AAA69257.1; -.  
 DR EMBL; AB000358; AAC75789.1; -.  
 DR PIR; G65055; G65055.  
 DR PDB; 1I52; 1I-JUL-01.  
 DR PDB; 1INI; 14-NOV-01.  
 DR PDB; 1INU; 11-JUL-01.  
 DR EcGene; EG1110; ISPD.  
 DR HAMAP; MF\_00108; -; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR TIGRFAMs; TIGR00453; ISPD; 1.  
 DR TRANSF; PIRSF006765; DPCME\_synch; 1.  
 KW Transferrase; Nucleotidyltransferase; Isoprene biosynthesis; Magnesium;  
 KW Manganese; Cobalt; 3D-structure; Complete proteome.  
 FT INIT MET 0  
 FT STRAND 7 13  
 FT HELIX 18 20  
 FT HELIX 26 28  
 FT STRAND 30 31  
 FT TURN 32 33  
 FT STRAND 34 35  
 FT HELIX 36 45  
 FT TURN 46 46  
 FT TURN 48 49

| SEQ    | SEQUENCE | 235 AA; | 25606 MW; | 2C090B31E2B61B6C CRC64; |
|--------|----------|---------|-----------|-------------------------|
| FT     | STRAND   | 50      | 57        |                         |
| FT     | TURN     | 59      | 60        |                         |
| FT     | HELIX    | 64      | 66        |                         |
| FT     | HELIX    | 68      | 71        |                         |
| TURN   | TURN     | 73      | 74        |                         |
| FT     | STRAND   | 75      | 79        |                         |
| FT     | HELIX    | 84      | 93        |                         |
| FT     | TURN     | 94      | 94        |                         |
| FT     | TURN     | 96      | 97        |                         |
| FT     | STRAND   | 100     | 103       |                         |
| TURN   | TURN     | 106     | 107       |                         |
| FT     | TURN     | 109     | 110       |                         |
| FT     | HELIX    | 113     | 120       |                         |
| FT     | HELIX    | 121     | 124       |                         |
| FT     | TURN     | 125     | 125       |                         |
| FT     | STRAND   | 130     | 135       |                         |
| FT     | STRAND   | 140     | 143       |                         |
| FT     | TURN     | 145     | 146       |                         |
| FT     | STRAND   | 150     | 154       |                         |
| FT     | TURN     | 157     | 158       |                         |
| FT     | STRAND   | 160     | 169       |                         |
| FT     | HELIX    | 170     | 182       |                         |
| TURN   | TURN     | 183     | 184       |                         |
| FT     | HELIX    | 190     | 196       |                         |
| FT     | TURN     | 197     | 198       |                         |
| STRAND | STRAND   | 202     | 205       |                         |
| TURN   | TURN     | 208     | 209       |                         |
| FT     | HELIX    | 216     | 226       |                         |

|                       |                  |                    |           |             |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match           | 43.4%;           | Score 520.5;       | DB 1;     | length 235; |
| Best Local Similarity | 49.8%;           | Pred. No. 2.3e-38; |           |             |
| Matches 11;           | Conservative 31; | Mismatches 72;     | Indels 9; | Gaps 4      |

QY AAVPAAGVGRKMADEKQYLLPLAGWTVEHTLFTLESDAPQKVAVALISEDPWEELS 68  
 Db AAVPAAGVGRKMQTSECKQYLLISGNTLIEHSVALHPKPKVVALISEGSRFQALP 68  
 QY IAAHPDITAPGGKEXADSVLSALKALEDIASENMVLVHDAARPCITGSDIHLDTLK 128  
 Db IAAHPDITVVDGGDEADSVLAKGLKAAAAGDAQWLVHDAARPCILHQBDEARILALSE 124  
 QY 129 NDVVGGLIALSHDTLKYVD--GDVTITATIDAKAVKRLALTQMKRYGMLDALQR--TEG 184  
 Db 125 TSSTGGILAAVVDITKRAEPKRNIAHTVPRNGIWEHALTQFFPRSEIHLCDITRALNEG 184  
 QY 185 NPAVTDEASALTELHGKPKIVEGRPDNKTIRSPEDLAAQPYM 227  
 Db 185 -ATITDEASALTECGFHPOVEGRPDNKTIRSPEDLAAEYLL 226

| AC | Q88471   | STANDARD | PRT | 236 AA |
|----|--|----------|-----|--------|
| DT | 28-FEB-2003 (Rel. 41, Created)                                       |          |     |        |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update)                          |          |     |        |
| DT | 15-SEP-2003 (Rel. 42, Last annotation update)                        |          |     |        |
| DE | 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60) |          |     |        |
| DE | (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP          |          |     |        |
| DE | cytidyllyltransferase) (MCT).  |          |     |        |
| GN | ISPD OR STY3055 OR T2831.  |          |     |        |
| OS | <i>Salmonella typhi</i> .  |          |     |        |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;    |          |     |        |
| OC | Enterobacteriaceae; <i>Salmonella</i> .                              |          |     |        |
| OX | NCBI_TaxID=601;  |          |     |        |

RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mangall K.L., Bentley S.D., Holden M.T.G., Sebalth M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Cornerton P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Ujvelcs K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Goara P., Parry C.,  
RA Quail M., Rutherford K., Simmonds W., Skelton J., Stevens K.,  
RA Whitehead S., Barrell B.G.,  
RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
RT enterica serovar Typhi CT18."  
RT Nature 413:848-852(2001).  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700991;  
RX MEDLINE=22513167; PubMed=12644504;  
RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyiadou V., Schwartz D.C., Blattner P.R.,  
RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2  
RT and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -I. FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-  
methy1-D-erythritol from CTP and 2C-methy1-D-erythritol 4-  
phosphate (by similarity).  
CC -I. CATALYTIC ACTIVITY: CTP + 2-C-methy1-D-erythritol 4-phosphate =  
diphosphate + 4-(cytidine 5'-diphospho)-2-C-methy1-D-erythritol.  
CC -I. PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
CC -I. SIMILARITY: BELONGS TO THE ISPD FAMILY.

CC -----  
 DR EXBL; AL627276; CAD06036.1; -.  
 DR EXBL; AE016843; AAO7038.1; -.  
 DR HAMAP; MF\_00108; -. 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR PIRSF; PIRSF006765; DPCME\_synth; 1.  
 DR TIGRFAMS; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 KW Transferase; Nucleotidyltransferase; Isoprene biosynthesis;  
 KW Complete proteome.  
 SQ SEQUENCE 236 AA; 25729 MW; DB4CB1315483A90E CRC64;

|                       |                 |                   |          |            |
|-----------------------|-----------------|-------------------|----------|------------|
| Query Match           | 43.0%           | Score 515.5       | DB 1     | Length 236 |
| Best Local Similarity | 49.8%           | Pred. No. 6.4e-38 |          |            |
| Matches 111           | Conservative 32 | Mismatches 71     | Indels 9 | Gaps 4     |

Qy 9 AAVPAAGVGKQADAPRQYLPAGKVTIETLRLESDAPQCVANALVEBPWPEIS 68  
 Db 10 AAVPAAGFRRRQGTCEPQYLSIGKLTLESHVALLAHPRVTVTVALISGCHRFQPLP 69  
 Qy 69 IAKHPDIIAPGCKERADSVLSATALEDIASENDWVLVMDNAPCTIGSDIHLDITLK 128  
 Db 70 IAHNPPIITVDDGNERADSVLAGLQAV---AKQWVLVMDNAPCHODLARLALISE 125  
 Qy 129 NDEVGIIALSSHDITLKAVD--GDTITATIDRKHYRALTIPQWFKYGMRLDALQR--TEG 184  
 Db 126 NSRVGGIILSPVROTMRKGEPEKNAIAHTVERADLMHALITPQFPEPRLHDLCTRALNEG 185  
 Qy 185 NPAVMDNANLFLSHKPKLYVGRDNIKLRPEDIALLAQRYM 227  
 Db 186 -ATTDEASALYCGHPALVGRADNKKVTRPEEDLALAEFTYL 227

| RESULT 10    | STANDARD;                         | PRT; | 236 AA. |
|--------------|-----------------------------------|------|---------|
| ISP.D. SALTY |                                   |      |         |
| ID           |                                   |      |         |
| ISP.D. SALTY |                                   |      |         |
| Q82M56.      |                                   |      |         |
| DT           |                                   |      |         |
| 28-FEB-2003  | (Rel. 41. Created)                |      |         |
| DT           |                                   |      |         |
| 28-FEB-2003  | (Rel. 41. Last sequence update)   |      |         |
| DT           |                                   |      |         |
| 28-FEB-2003  | (Rel. 41. Last annotation update) |      |         |



DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP  
 DE cytidyltransferase) (MCT).  
 GN ISPd OR STM2930.  
 OC Salmonella typhimurium.  
 OC Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OK NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=1167609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856 (2001).  
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidylyl-2C-  
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-  
 CC phosphate (by similarity).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPd FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE008833; AAL21810.1; -.  
 DR StGene; SG77777; ISPd.  
 DR HAMAP: MF\_00108; -; 1.  
 DR InterPro: IPR001228; ISPd\_synthase.  
 DR Pfam: PF01128; ISPd; 1.  
 DR PIRSF: PIRSF006765; DPcME\_synth; 1.  
 DR TIGRFAMs: TIGR00453; ISPd; 1.  
 DR PROSITE: PS01295; ISPd; 1.  
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;  
 DR Complete proteome.  
 KW Complete proteome.  
 SQ SEQUENCE 236 AA; 25755 MW; C4E631315492B80B CRC64;  
 Query Match 42.8%; Score 513.5; DB 1; Length 236;  
 Best Local Similarity 49.8%; Pred. No. 9,6e-38;  
 Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;  
 QY 9 AVPAAGVGKQADRPKQYLPAGKTVIEHTLRLSLSDAFQKAVAVASVEDPYWPELS 68  
 DB 10 AVPAAGFGRMRQTCPKQYLSIGKTLIESSVHLLHMPRTVRVIAISPDHFAALP 69  
 QY 69 IAKHPDITTAPEGKRAVSVALKALIEDIASENDWVYVHDAARCLVSGDHLQIDTLK 128  
 DB 70 IANHQITVVDGNGRARSVALGLQAV-----AKAWVAVHDAARCLVQDDIARLALISE 125  
 QY 129 NDPVGGIILASDHLKAVD-GDTITATIDRKHWARALTPOMFKYGMRLDALQR--TEG 184  
 DB 126 NERVGGIILASPDVDMKREPERKNAIAHVERADLMHNLTPFPFRELIDLTALNMG 185  
 QY 185 NPAVTDEASALELLHKEKTVIEGRPDNKTITREPDLAAQPTM 227  
 DB 186 -ATITDEASALEYCGFHPALVGEGRADNIVKTRPEDIALAEFYL 227  
 RESULT 11  
 ID ISPd\_YERPE STANDARD; PRT; 241 AA.  
 AC Q8ZBP6;

DT 23-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 23-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP  
 DE cytidyltransferase) (MCT).  
 GN ISPd OR YPO3361 OR Y0828.  
 OC Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OK NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Tiltball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher G., Mungall K.L.,  
 RA Baker S., Baaham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";  
 RL Nature 413:523-527 (2001).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RC STRAIN=KIMS / Biovar Mediaevalis;  
 RX MEDLINE=22137863; PubMed=12142430;  
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,  
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
 RA Bertherton J.D., Lindler L.B., Brubaker R.R., Plano G.V.,  
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,  
 RA Perry R.D.;  
 RT "Genome sequence of Yersinia pestis KIM.";  
 RL J. Bacteriol. 184:4601-4611 (2002).  
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidylyl-2C-  
 CC methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-  
 CC phosphate (by similarity).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPd FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ414156; CAC92591.1; -.  
 DR EMBL: AE013685; AAM64413.1; -.  
 DR PIR: AC0408; ACO408.  
 DR HAMAP: MF\_00108; -; 1.  
 DR InterPro: IPR001228; ISPd\_synthase.  
 DR Pfam: PF01128; ISPd; 1.  
 DR PIRSF: PIRSF006765; DPcME\_synth; 1.  
 DR TIGRFAMs: TIGR00453; ISPd; 1.  
 DR PROSITE: PS01295; ISPd; 1.  
 DR Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;  
 DR Complete proteome.  
 KW Complete proteome.  
 SQ SEQUENCE 241 AA; 26401 MW; EF4CA6846BDAC261 CRC64;  
 Query Match 42.7%; Score 512.5; DB 1; Length 241;  
 Best Local Similarity 49.1%; Pred. No. 1.2e-37;  
 Matches 111; Conservative 35; Mismatches 71; Indels 9; Gaps 4;  
 QY 9 AVPAAGVGKQADRPKQYLPAGKTVIEHTLRLSLSDAFQKAVAVASVEDPYWPELS 68  
 DB 13 AVPAAGISRLVDCPKQYLVGKTLIEHAIISLHRRIRGRVIVIHQDQFSPRLS 72  
 QY 69 IAKHPDITTAPEGKRAVSVALKALIEDIASENDWVYVHDAARCLVSGDHLQIDTLK 128

Db 73 VADPRLSTVGGDQGANSMAGLQ---LAGAENVIVHADARPCHLDDLSRLISTE 128  
 QY 129 NDEVGIIALSSDITLKHVDG--DTTATIDRCKHWRALTPQKFKGTLDALQRT--EG 184  
 Db 129 CQVGGIIILAPVDIMKRRAPGIAIAFTVDRQDLWALPOLFPLELTKLSRALREG 188  
 QY 185 NPATVEASALELGHKPKIVEGRPNIKITRPEDLALAQFYEQQ 230  
 Db 189 -VAITDEASALEHCGHPIILVTRSDNIKITRPEDLALAFYLTOR 233

RESULT 12  
 ISPD\_PASMU STANDARD; PRT; 238 AA.

ID ISPD\_PASMU STANDARD; PRT; 238 AA.  
 AC P57953;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP  
 cytidyltransferase) (MCT).  
 GN ISPD OR PM1608.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 OX NCBI\_TaxId=747;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70.  
 RX MEDLINE=2115866; PubMed=11248100;  
 RA May 5.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).  
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-  
 METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-  
 PHOSPHATE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nomenclature: terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL, AB006198; AK03692.1; -  
 DR HSSP; Q46893; I152.  
 DR HAMAP; MF\_00108; -; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR PIRSF; PIRSF006765; DPCME synth; 1.  
 DR TIGRFAMs; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR TRANSFERASE; Nucleotidyltransferase; Isoprene biosynthesis;  
 KW Complete proteome.  
 SQ SEQUENCE 238 AA; 26195 MW; 6AP21525CB4CFE22 CRC64;

Query Match 42.1%; Score 505; DB 1; Length 238;  
 Best Local Similarity 47.6%; Pred. No. 5.4e-37;  
 Matches 107; Conservative 33; Mismatches 79; Indels 6; Gaps 4;

QY 9 AVPAAGVGKMQADRPKQYIPLAGKTVIEHTLTLLSDAFQKVAVAISVEDPYPELS 68  
 Db 14 AVPAAGIGSRMQMDPKQYIHLHGTIIIEHTLSVLLGYPILTEKILAAVANDPYISTCP 73  
 QY 69 IAKHPDITAPGGERADSVLSALKALEDA--SEWDWLVYHDAAPPCLTGSDIHQIDFLK 126  
 Db 74 LITHPKIQVLEGGSSADSVLNGINLVAKSAVONSEDFWVAVHDAAPPCLTGDDKLVGV 133

QY 127 LKNDEVGIIALSSDITLKH--VDGTTATIDRCKHWRALTPQKFKGTLDALQRTEN 185  
 Db 124 --EDNKGAILIAPDITLKRALHNOQHITRDSQMLAQTPFPPIATTLQAALQALQA 191  
 QY 186 P-AVDEASALELGHKPKIVEGRPNIKITRPEDLALAQFYEQQ 229  
 Db 192 GLQVITDEASAMEFAGFRPHLVAGSDNIKITRPEDFALAEFLYSR 236

RESULT 13  
 ISPD\_BUCAL STANDARD; PRT; 237 AA.

ID ISPD\_BUCAL STANDARD; PRT; 237 AA.  
 AC P57455;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase) (MEP  
 cytidyltransferase) (MCT).  
 GN ISPD OR BU420.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OX NCBI\_TaxId=118099;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998.  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS.";  
 RL Nature 407:81-86 (2000).  
 CC -1- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-  
 METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-  
 PHOSPHATE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nomenclature: terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL, AP003119; BAB13118.1; -  
 DR HSSP; Q46893; I152.  
 DR HAMAP; MF\_00108; -; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR PIRSF; PIRSF006765; DPCME synth; 1.  
 DR TIGRFAMs; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR TRANSFERASE; Nucleotidyltransferase; Isoprene biosynthesis;  
 KW Complete proteome.  
 SQ SEQUENCE 237 AA; 26922 MW; 388AFDEP35BCB706 CRC64;

Query Match 39.6%; Score 474.5; DB 1; Length 237;  
 Best Local Similarity 44.4%; Pred. No. 2.5e-34;  
 Matches 99; Conservative 50; Mismatches 67; Indels 7; Gaps 3;

QY 9 AVPAAGVGKMQADRPKQYIPLAGKTVIEHTLTLLSDAFQKVAVAISVEDPYPELS 68  
 Db 13 AVPAAGIGSRMQMDPKQYIHLHGTIIIEHTLSVLLGYPILTEKILAAVANDPYISTCP 72  
 QY 69 IAKHPDITAPGGERADSVLSALKALEDA--SEWDWLVYHDAAPPCLTGSDIHQIDFLK 126  
 Db 73 ISSNRITISVVGGERKINSVLSGLIVYKVV---DWIVYHDAVAPCLSYDLEKLSITIK 128

OY 129 NDPVGGIILASHDITKH--VDGDTITATIDKHWRALTPQMKMLRDALQK-TEGN 185  
 DB 129 KNPVGAIIIRPVSDITKISNLKQKAVITVKKMHWALTPQKLVKLNCKLTKITDOQ 188  
 OY 186 PAVTDASALIELGHEKPKIVEGRPNIKITRPEDIALAOFYME 228  
 DB 189 ISVTDBASALEYCGVNPFLVLGSCRNKIKITWPEDLVLANFYLK 231

## RESULT 14

ISPD\_WIGBR STANDARD; PRT; 229 AA.

AC Q8D23;  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP)  
 DE cytidyltransferase) (MCT).  
 OS Wigglesworthia glossinidia brevipalpis.  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wigglesworthia.  
 NCBI\_TaxID=36870;  
 [1]  
 SEQUENCE FROM N.A.  
 RX Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 Aksoy S.,  
 "Genome sequence of the endocellular obligate symbiont of tsetse  
 flies, Wigglesworthia glossinidia.";  
 Nat. Genet. 32:402-407(2002).  
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-  
 methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-  
 phosphate (By similarity).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB063522; BAC24678.1; --  
 DR HAMAP; MF\_00108; --; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR TIGRFAMs; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR TRANSFERASE; Nucleotidyltransferase; Isoprene biosynthesis;  
 KM Complete proteome.  
 SQ SEQUENCE 229 AA; 26240 MW; 9CA1F91717B89FD1 CRC64;

Query Match 38.6%; Score 462.5; DB 1; Length 229;  
 Best Local Similarity 44.3%; Pred. No. 2.6e-33;  
 Matches 97; Conservative 43; Mismatches 74; Indels 5; Gaps 2;

OY 10 VVPAAGVGKRMQADPRKQYLPAGKIVIEHTITRLLESDAFOKAVASVEDPYPELST 69  
 DB 10 IFFPAGISGRKQYKPKQYIKIKNTTIEHSISLFDIKIYVKIILAIKNDYWNFKLSI 69  
 OY 70 AKGPDIITAPGSKRADSVLSALKALIEDIASENDVVLVHDAARPCLTGSDIHLCITDTKN 129  
 DB 70 LKKNKINIVIGKSTESVISLTK---FVSKVAVVLVHDAVRPCGLHNDKLNKLKLVINI 125  
 OY 130 DPVGGIILASHDITKHVDGDTITATIDKHWRALTPQMKMLRDALQK-TEGNPAV 188

DB 126 SPFGAIIAIPYIDYVKKSYGNFISHTIKRNMALRTPQKLVNKLNLCKLITSKGBII 185  
 OY 189 TDBASALELGHKPKIVEGRPNIKITRPEDIALAOFYME 227  
 DB 186 TDBSASALEKCGYKLVHGRSDNIKITYPEDIANPFFI 224

## RESULT 15

ISPD\_HUGAP STANDARD; PRT; 236 AA.

AC Q8K9D6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60)  
 DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP)  
 DE cytidyltransferase) (MCT).  
 GN ISPD OR BUSG405.  
 OS Buchnera aphidicola (subsp. Schizaphis graminum).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 NCBI\_TaxID=98794;  
 [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=22084549; PubMed=12089438;  
 RX Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,  
 Verneegren J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;  
 "50 million years of genomic stasis in endosymbiotic bacteria.";  
 Science 296:2376-2379(2002).  
 CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-  
 methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-  
 phosphate (By similarity).  
 CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =  
 diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.  
 CC -1- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; third step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB014116; AAM67955.1; --  
 DR HAMAP; MF\_00108; --; 1.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR PIRSF; PIRSF006765; DPCMG\_synth; 1.  
 DR TIGRFAMs; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR TRANSFERASE; Nucleotidyltransferase; Isoprene biosynthesis;  
 KM Complete proteome.  
 SQ SEQUENCE 236 AA; 26967 MW; B3CCB1F88740DD25 CRC64;

Query Match 38.2%; Score 457.5; DB 1; Length 236;  
 Best Local Similarity 43.3%; Pred. No. 7.5e-33;  
 Matches 100; Conservative 42; Mismatches 82; Indels 7; Gaps 3;

OY 2 NPTIOCAVAVPAGVGKRMQADPRKQYLPAGKIVIEHTITRLLESDAFOKAVASVED 61  
 DB 6 SPKLIATLIVPAGISGRKMDPRKQYIKIKDCTIIEYTLKTLSSHNIVRIYVLSHQED 65  
 OY 62 PYWPELSIAHPDITITAPGSKRADSVLSALKALIEDIASENDVVLVHDAARPCLTGSDIH 121  
 DB 66 NFFPKLSSISDILRFVSTLGNRERHSVLSGL---ITTPAKVVIHDAVAPCLSYDLE 121  
 OY 122 LQDITLKNDDVGGIILASHDITKHVD--GDITITATIDKHWRALTPQMKMLRDAL 179  
 DB 122 NLAIITNTVGGIILAPVCDITIKSRKNTLITHTIPKQMLHALLTPQKLPINLAFCL 181  
 OY 180 QR-TEGNPAVTDASALELGHKPKIVEGRPNIKITRPEDIALAOFYMEQ 229

Thu Jan 29 17:41:18 2004

us-09-941-947a-10.rsp

Page 10

Db 182 KXIVEDKXNITDPAASHLSTGTHPLIVGXYXNIXITVPEDLIFAFIIXE 232

Search completed: January 29, 2004, 15:50:40  
Job time : 5.34994 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:14 ; Search time 17.2185 Seconds  
(without alignments)  
3461.979 Million cell updates/sec

Title: US-09-941-947a-10

Perfect score: 1199

Sequence: 1 NMPTICWAVVPAAGVGRM.....IKTRPEDIALAQFMEQQA 231

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 605   | 50.5        | 237    | 16    | Q8DC60      |
| 2          | 520.5 | 43.4        | 236    | 16    | Q8FE05      |
| 3          | 492   | 41.0        | 249    | 16    | Q8EBR2      |
| 4          | 462.5 | 38.6        | 229    | 16    | Q8D223      |
| 5          | 301   | 25.1        | 132    | 2     | Q8EYV4      |
| 6          | 283.5 | 23.6        | 302    | 10    | Q64726      |
| 7          | 272.5 | 22.7        | 234    | 16    | Q8D151      |
| 8          | 272   | 22.7        | 474    | 2     | Q48230      |
| 9          | 258   | 21.5        | 474    | 2     | Q48154      |
| 10         | 256   | 21.4        | 238    | 16    | Q8F7A0      |
| 11         | 249.5 | 20.8        | 451    | 16    | Q8G0H4      |
| 12         | 248.5 | 20.7        | 237    | 2     | Q8BK19      |
| 13         | 248.5 | 20.7        | 248    | 16    | Q8FMT3      |
| 14         | 246   | 20.5        | 236    | 16    | Q8Y832      |
| 15         | 232.5 | 19.4        | 246    | 16    | Q8KJ33      |
| 16         | 229.5 | 19.1        | 240    | 16    | Q8DY07      |

|    |       |      |     |    |        |                    |
|----|-------|------|-----|----|--------|--------------------|
| 17 | 228.5 | 19.1 | 239 | 16 | Q8E4B4 | Q8E4B4 streptococc |
| 18 | 227   | 18.9 | 237 | 16 | Q92CV0 | Q92CV0 listeria in |
| 19 | 224.5 | 18.7 | 291 | 16 | Q8G7B2 | Q8G7B2 bifidobacte |
| 20 | 218.5 | 18.2 | 238 | 16 | Q8CQ77 | Q8CQ77 staphylococ |
| 21 | 214.5 | 17.9 | 238 | 16 | Q99WM8 | Q99WM8 staphylococ |
| 22 | 209   | 17.4 | 235 | 16 | Q970E5 | Q970E5 streptococc |
| 23 | 209   | 17.4 | 235 | 16 | Q8DP12 | Q8DP12 streptococc |
| 24 | 200.5 | 16.7 | 238 | 16 | Q99WX2 | Q99WX2 staphylococ |
| 25 | 200.5 | 16.7 | 240 | 16 | Q8EWC7 | Q8EWC7 mycoplasma  |
| 26 | 196.5 | 16.4 | 238 | 16 | Q8N1T0 | Q8N1T0 staphylococ |
| 27 | 171.5 | 14.3 | 241 | 2  | Q8GPF2 | Q8GPF2 streptococ  |
| 28 | 170   | 14.2 | 324 | 11 | Q8CAE0 | Q8CAE0 mus musculu |
| 29 | 141   | 11.8 | 229 | 17 | Q58136 | Q58136 pyrococcus  |
| 30 | 133   | 11.1 | 203 | 10 | Q8S0P7 | Q8S0P7 oryza sativ |
| 31 | 133   | 11.1 | 362 | 11 | Q8BR14 | Q8BR14 mus musculu |
| 32 | 125.5 | 10.5 | 246 | 2  | Q9XB85 | Q9XB85 amycolatops |
| 33 | 122   | 10.2 | 734 | 5  | Q81273 | Q81273 plasmodium  |
| 34 | 118   | 9.8  | 458 | 16 | Q9CEP8 | Q9CEP8 lactococcus |
| 35 | 113   | 9.4  | 187 | 11 | Q8C934 | Q8C934 mus musculu |
| 36 | 112   | 9.3  | 460 | 16 | Q8G5P1 | Q8G5P1 bifidobacte |
| 37 | 109   | 9.1  | 492 | 16 | Q9CD44 | Q9CD44 mycobacteri |
| 38 | 109   | 9.1  | 513 | 16 | P96382 | P96382 mycobacteri |
| 39 | 108.5 | 9.0  | 295 | 2  | Q9X364 | Q9X364 bacillus an |
| 40 | 107.5 | 9.0  | 242 | 16 | Q98CM6 | Q98CM6 rhizobium 1 |
| 41 | 106.5 | 8.9  | 456 | 16 | Q8ZS87 | Q8ZS87 yersinia pe |
| 42 | 106.5 | 8.9  | 458 | 16 | Q8CZF5 | Q8CZF5 yersinia pe |
| 43 | 105.5 | 8.8  | 291 | 17 | Q9UZ17 | Q9UZ17 pyrococcus  |
| 44 | 103.5 | 8.6  | 482 | 16 | Q8CJX6 | Q8CJX6 streptomyce |
| 45 | 102   | 8.5  | 239 | 16 | Q8FXZ3 | Q8FXZ3 brucella su |

## ALIGNMENTS

### RESULT 1

| ID | Q8DC60  | PRELIMINARY;                | PRT; | 237 AA. |
|----|---|-----------------------------|------|---------|
| AC | Q8DC60;   |                             |      |         |
| DT | 01-MAR-2003 (TREMBL)  | 23, Created)                |      |         |
| DT | 01-MAR-2003 (TREMBL)  | 23, Last sequence update)   |      |         |
| DT | 01-MAR-2003 (TREMBL)  | 23, Last annotation update) |      |         |
| DE | 4-diphosphocytidylyl-2-methyl-D-erithritol synthase.            |                             |      |         |
| GN | VY1582.   |                             |      |         |
| OS | Vibrio vulnificus.  |                             |      |         |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;     |                             |      |         |
| OC | Vibrionaceae; Vibrio.   |                             |      |         |
| OX | NCBI_TaxID=672;   |                             |      |         |
| RN | [1]   |                             |      |         |
| RP | SEQUENCE FROM N.A.  |                             |      |         |
| RC | STRAIN=CMCP6;   |                             |      |         |
| RA | Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H., |                             |      |         |
| RA | Choy H.E.;  |                             |      |         |
| RT | "Complete genome sequence of Vibrio vulnificus CMCP6."          |                             |      |         |
| RL | Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.         |                             |      |         |
| DR | EMBL; AB016802; AA010005.1; -                                   |                             |      |         |
| KW | Complete proteome.  |                             |      |         |
| SC | SEQUENCE 237 AA; 26005 MW; 35D0952D12B47ECD CRC64;              |                             |      |         |

|    |     |  |     |
|----|-----|--|-----|
| QY | 9   | AYVPAAGVGRMADREKQVPLAGKTVIHTLTLLLESDAPQVAVASVDEPYNELS    | 68  |
| DB | 10  | AYVPAAGVGRMADREKQVPLAGKTVIHTLTLLLESDAPQVAVASVDEPYNELS    | 69  |
| QY | 69  | IAPKPDITAPGKERADSVLSAKLEDIASERDWTVAAPPCULGSDIHQIDTLK     | 128 |
| DB | 70  | IAPKPDITAPGKERADSVLSAKLEDIASERDWTVAAPPCULGSDIHQIDTLK     | 129 |
| QY | 129 | NDPVGGLIASSHDTLAGVNG-DITITIRKAVRALLTQMPKRYGLRDL-QRTGNP   | 186 |
| DB | 130 | SHETGILATPVRDITKRRANAQWIDHTVDRNALMHALTQMPKRYGLRDL-QRTGNP | 189 |

QY 187 AVTDEASALELGHKPKIVEGRPDNITKTRPEDLALAQFYMQ 230  
 DB 190 AITDEASALEMGRBEPALVQSCSSNITKVPEDLALAEVYSRE 233

## RESULT 2

Q8E8T5 PRELIMINARY; PRT; 236 AA.

AC Q8E8T5; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (EC 2.7.7.60).  
 GN YGAP OR C3314.

OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;

SEQUENCE FROM N.A.  
 RC STRAIN-06.H1 / CPT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasco D., Buckles B.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner P.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 DR EMBL; AB016765; AAN01763.1; -;  
 KW Nucleotidyltransferase; Transferase; Complete proteome.

SEQUENCE 236 AA; 25779 MW; E2065C781E122202 CRC64;

Query Match 43.4%; Score 520.5; DB 16; Length 236;

Best Local Similarity 49.8%; Pred. No. 1.6e-36; Mismatches 71; Indels 9; Gaps 4;

Matches 111; Conservative 32; Mismatches 71; Indels 9; Gaps 4;

QY 9 AVVPAAGVGRMQRADPRKQYLPPLAGKTVIEHTLRLLESDAFQKAVAVISVEDPYMPELS 68

DB 10 AVVPAAGVGRMQRADPRKQYLPPLAGKTVIEHTLRLLESDAFQKAVAVISVEDPYMPELS 69

QY 69 IAKHPDITTAPOGKERADSVLSALKLEDIASENDVAVHDAARPCLTGSDIHQIDTLK 128

DB 70 IAKHPDITTAPOGKERADSVLSALKLEDIASENDVAVHDAARPCLTGSDIHQIDTLK 125

QY 129 NDPVGGILASSHDLTKHVG-DITATIDRKIVWRALTPQMFKYGMILRALQ--TEG 184

DB 126 TERTGIIAAPPYRDTMKRAEPGRKALARTVDKNGIMHAIPTQFFRELLHDCITLALNG 185

QY 185 NPAVTDEASALELGHKPKIVEGRPDNITKTRPEDLALAQFYM 227

DB 186 AITDEASALEMGRBEPALVQSCSSNITKVPEDLALAEVYSRE 227

RESULT 3

Q8E8R2 PRELIMINARY; PRT; 249 AA.

AC Q8E8R2; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE 4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.  
 GN ISPD OR S03438.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;

SEQUENCE FROM N.A.

RC STRAIN=NR-1;  
 RX MEDLINE=22297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos B.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seebach R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer R., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.U., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umeyama L.A., White O., Wolf A.M.,  
 RA Vamathevan U., Weidman J., Imprim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AB015780; AAN56435.1; -;  
 RX TIGR; S03438; -;  
 KW Complete proteome.

SEQUENCE 249 AA; 26590 MW; 5B8BE0D57590A49 CRC64;

Query Match 41.0%; Score 492; DB 16; Length 249;

Best Local Similarity 48.0%; Pred. No. 4.7e-34; Mismatches 106; Conservative 33; Mismatches 76; Indels 6; Gaps 3;

QY 9 AVVPAAGVGRMQRADPRKQYLPPLAGKTVIEHTLRLLESDAFQKAVAVISVEDPYMPELS 68

DB 28 AVVPAAGVGRMQRADPRKQYLPPLAGKTVIEHTLRLLESDAFQKAVAVISVEDPYMPELS 87

QY 69 IAKHPDITTAPOGKERADSVLSALKLEDIASENDVAVHDAARPCLTGSDIHQIDTLK 128

DB 88 IAKHPDITTAPOGKERADSVLSALKLEDIASENDVAVHDAARPCLTGSDIHQIDTLK 143

QY 129 NDPVGGILASSHDLTKHVG-DITATIDRKIVWRALTPQMFKYGMILRALQ--RTGNP 186

DB 144 QPFGAIIAMPVRRITMKRASLSGINSIVCRDWMHAIPTQFFRELLHDCITLALNG 203

QY 187 AVTDEASALELGHKPKIVEGRPDNITKTRPEDLALAQFYM 227

DB 204 VITDEASALEMGRBEPALVQSCSSNITKVPEDLALAEVYSRE 244

RESULT 4

Q8D223 PRELIMINARY; PRT; 229 AA.

AC Q8D223; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE YGP protein.  
 GN YGP.

OS Wigglesworthia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wigglesworthia.  
 OX NCBI\_TaxID=164609;

SEQUENCE FROM N.A.

RC MEDLINE=22297718; PubMed=12219091;

RA Akman L., Yamashta A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 RA Aksoy S.;

RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 flies, Wigglesworthia glossinidia.";

RL Nat. Genet. 32:402-407(2002).  
 RX EMBL; AB063522; BAC24678.1; -;

KW Complete proteome.

SEQUENCE 229 AA; 26240 MW; 9CA1F49717B9FD1 CRC64;

Query Match 38.6%; Score 462.5; DB 16; Length 229;

Best Local Similarity 44.3%; Pred. No. 1.4e-31; Mismatches 97; Conservative 43; Mismatches 74; Indels 5; Gaps 2;

QY 10 VPAAGVGRMQRADPRKQYLPPLAGKTVIEHTLRLLESDAFQKAVAVISVEDPYMPELS 69

DB 10 IFAAGVGRMQRADPRKQYLPPLAGKTVIEHTLRLLESDAFQKAVAVISVEDPYMPELS 69

QY 70 AKHPDITTAPOGKERADSVLSALKLEDIASENDVAVHDAARPCLTGSDIHQIDTLK 129

DB 70 AKHPDITTAPOGKERADSVLSALKLEDIASENDVAVHDAARPCLTGSDIHQIDTLK 125

QY 130 DPVGGILATSSHTLKHVGDGTITATDRKHWRALTPQWKYGMALDQ-RTBGRNAV 188  
 DB 126 SPFGALIAPIYIPVKSYPNFSHTIKRKLTPALPOLKILNCKLITTSKGEII 185  
 QY 189 TDEASALELGHKPKIVEGRPDNIIKTRPEDLALAOFPYM 227  
 DB 186 TDESSALEKCGYKLNLYHGSNDNIIKITYPEDLNPANPFI 224

RESULT 5  
 QY09EY4 PRELIMINARY; PRT; 132 AA.  
 ID 09EY4  
 AC 09EY4  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE YBP (Fragment).  
 GN YBP.  
 OS Klebsiella aerogenes.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=28451.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N70;  
 RX MEDLINE=20566700; Pubmed=11114933;  
 RA Kolko M.M., Kapetanovich L.A., Lawrence J.G.;  
 RT "Alternative pathways for streptococcal synthesis in Klebsiella  
 aerogenes.";  
 RT Bacteriol. 183:328-335 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N70;  
 RA Seifstein T.A., Lawrence J.G.;  
 RT "Methionine recycling in Klebsiella aerogenes.";  
 RT Bacteriol. 0:0-0 (2001).  
 DR EMBL; AF308468; AAC42460.1; -.  
 DR HSSP; Q46893; 1152.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR NON\_TER 132 132  
 SQ SEQUENCE 132 AA; 14033 MW; 69P919A5084B2AFB CRC64;

Query Match 25.1%; Score 301; DB 2; Length 132;  
 Best Local Similarity 48.8%; Pred. No. 4.1e-18;  
 Matches 62; Conservative 18; Mismatches 43; Indels 4; Gaps 1;

QY 9 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLTLRLSDAQKVAIVSDEDPYRPELS 68  
 DB 10 AVVPAAGVGRKQADRPKQYLPAGKTVIEHTLTLRLSDAQKVAIVSDEDPYRPELS 69  
 QY IAKHPDITAPGKGRADSVLSALKALIEDIASENDVVLVHDARPCLTSDIHLAQIDTLK 128  
 DB 70 IAKHPDITAPGKGRADSVLSALKALIEDIASENDVVLVHDARPCLTSDIHLAQIDTLK 125  
 QY 129 NDVPGGI 135  
 DB 126 TSVVGGI 132

RESULT 6  
 ID 064726 PRELIMINARY; PRT; 302 AA.  
 AC 064726; Q9LJ91;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Putative sugar nucleotide phosphorylase  
 DE (4-diphosphocytidylyl-2C-methyl-D-erythritol synthase)  
 DE (2-C-methyl-D-erythritol 4-phosphate cytidyltransferase).  
 GN ATG302500 OR ISPD OR ATWEPCT OR ATG20500/T8K22.20.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X., Kaul S., Tomlin C.D., Benito M.-I., Creeasy T.H., Haas B.J.,  
 Wu D., Walch R., Rongling C.M., Koo H., Fujii C.Y., Uterback T.R.,  
 Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RT Arabidopsis thaliana chromosome 2 BAC T8K22 genomic sequence."  
 RT Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20300921; Pubmed=10841550;  
 RA Rohdich F., Wungstentewekul J., Eisenreich W., Richter G.,  
 Schuhr C.A., Hecht S., Zenk M.H., Bacher A.;  
 RT Biosynthesis of terpenoids. 4-diphosphocytidylyl-2C-methyl-D-erythritol  
 RT synthase of Arabidopsis thaliana."  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:6451-6456 (2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RA Okada K., Kawade H., Kuzuyama T., Takagi M., Seto H., Kamiya Y.;  
 RT "2-C-methyl-D-erythritol 4-phosphate cytidyltransferase."  
 RT Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 Nakajima M., Enji A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 Hayashizaki Y., Shinozaki K.;  
 RT Arabidopsis thaliana full-length cDNA."  
 RT Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AC004136; AAC18936.2; -.  
 DR EMBL; AF230737; AAF61714.1; -.  
 DR EMBL; AB037876; BAB21592.1; -.  
 DR EMBL; AK18110; BAC42737.1; -.  
 DR HSSP; Q46893; 1152.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR TRFAMS; TIGR00453; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 DR TRANSFERASE.  
 SQ SEQUENCE 302 AA; 33937 MW; 7881DC5C8C37B06 CRC64;

Query Match 23.6%; Score 283.5; DB 10; Length 302;  
 Best Local Similarity 34.2%; Pred. No. 4e-16;  
 Matches 80; Conservative 46; Mismatches 83; Indels 25; Gaps 10;

QY 10 VVPAAGVGRKQADRPKQYLPAGKTVIEHTLTLRLSDAQKVAIVSDEDPYRPELS 63  
 DB 82 ILLAGGQGRKQADRPKQYLPAGKTVIEHTLTLRLSDAQKVAIVSDEDPYRPELS 135  
 QY 64 WPELSTAKHPDITAPGKGRADSVLSALKALIEDIASENDVVLVHDARPCLTSDIHL 122  
 DB 136 IFEYBESIDVDLRFALPKGRQDSYVSGLOEI-DVNSE--LVCHDSARPLWTEDE- 191  
 QY 123 QIDTLKNDPVGG--ILALSSHTLKHVGDGTITATDRKHWRALTPQWKYGMALDQ-RTBGRNAV 178  
 DB 192 --KVLKGSNAVGAIVLGPAPAKATIKVNSDSLVTKTIDRKLTMEMQFQVIRPELKKGF 249  
 QY 179 -LQRTGKPAVTDASALELGHKPKIVEGRPDNIIKTRPEDLALAOFPYM 231  
 DB 250 ELVNSG-LEVTLDVSVIVYIKHPVYVSGSYTNIKVTTPDDLILAEIRLSDS 302

RESULT 7  
 ID 08D191 PRELIMINARY; PRT; 234 AA.  
 AC 08D191;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)



DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
 DE 4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.  
 GN TIR0605.  
 OS Synecococcus elongatus (Thermosynechococcus elongatus).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
 OX NCBI\_TaxID=32046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BP-1;  
 RX MEDLINE=22225144; PubMed=12240634;  
 RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
 RA Watanabe A., Itiguchi M., Kawashima K., Kimura T., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsuno M., Matsuno A., Nakazaki N.,  
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the thermophilic cyanobacterium  
 RT Thermosynechococcus elongatus BP-1."  
 RL DNA Res. 9:123-130(2002).  
 DR EMBL; AP05371; BAC08157.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 234 AA; 25592 MW; EEBFIDBB32BC18E CRC64;

Query Match 22.7%; Score 272.5; DB 16; Length 234;  
 Best Local Similarity 35.1%; Pred. No. 2.5e-15;  
 Matches 80; Conservative 38; Mismatches 97; Indels 13; Gaps 7;  
 QY 10 VVPAAGGKRMQADPRQYLPAGKTVLEHTLRLLESDAPQKVAVAISVED-PYM--- 64  
 DB 4 LIPPAAGGKRMGSHNKLRLQLGKPLAWTLAAVAALIMWIGVIGQPEDFPIEMALL 63  
 QY 65 PELSLAKHEDITTAGGKERADSVLSALKALEDIASNDVTVVHDAAPCLTGSIDHQT 124  
 DB 64 EELNLRQPVHLLT-GEETQASVPHGLQALPKTAEQ--VLIHGGR-CLATPPLINRC 117  
 QY 125 DELKNDPVGGILALSSHDTLKNVGDG-TTATIDRKHWRALTPOMFKYGMRLDQRT 183  
 DB 118 AQAAGTVAAGLIAVAPVQKTIKIVNEGAVVQTPREDSLMAQTPGCFVNEPLRIHMAV 177  
 QY 184 GNP-AVTDASALELIGHEKPKIVBGPDKIKITRPEDLALAQFYEQ 230  
 DB 178 AKGWETDDAALFERLGVAVHIVLGEETLKITTPSDIPLAERILQHR 225

RESULT 8  
 Q48230 PRELIMINARY; PRT; 474 AA.  
 ID Q48230  
 AC Q48230;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
 DE Bifunctional ribulose 5-phosphate reductase/CDP-ribitol  
 DE pyrophosphorylase.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RM135;  
 RX MEDLINE=95272382; PubMed=7752865;  
 RA van Eldere J., Brophy L., Loynds B., Celis P., Kroil J.S., Moxon E.R.,  
 RA Hancock I., Carman S.;  
 RT "Region II of Haemophilus influenzae type b capsulation locus is  
 RT involved in serotype-specific polysaccharide synthesis."  
 RL Mol. Microbiol. 15:107-118(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RM135;  
 RX MEDLINE=99194706; PubMed=10094675;  
 RA Follens A., Veiga-da-Cunha M., Merckx R., an Schaftingen E.,  
 RA van Eldere J.;  
 RT "Accl of Haemophilus influenzae type a Capsulation Locus Region II  
 RT encodes a bifunctional ribulose 5-phosphate reductase-CDP-ribitol

RT pyrophosphorylase."  
 RL J. Bacteriol. 181:2001-2007(1999).  
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASIS/REDUCTASES  
 CC (SPR) FAMILY.  
 DR EMBL; X78559; CA555303.1; -.  
 DR InterPro; IPR002198; ADH\_short.  
 DR InterPro; IPR001228; ISPD synthase.  
 DR Pfam; PF00106; adh\_short\_1.  
 DR Pfam; PF01128; ISPD\_1.  
 DR PROSITE; PS00061; ADH\_SHORT; 1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 474 AA; 52467 MW; DC42A2B245318A95 CRC64;

Query Match 22.7%; Score 272; DB 2; Length 474;  
 Best Local Similarity 34.6%; Pred. No. 7.3e-15;  
 Matches 81; Conservative 34; Mismatches 85; Indels 34; Gaps 9;  
 QY 10 VVPAAGGKRMQADPRQYLPAGKTVLEHTLRLLESDAPQKVAVAISVED-PYM 63  
 DB 9 IILAGVSRMGICYPQPSFKLAKTALHTLALFOHKEIDEIIVSERTSYRIED-I 67  
 QY 64 MPELSIAKHEDITTAGGKERADSVLSALKALEDIASNDVTVVHDAAPCLTGSIDH 114  
 DB 68 VSKLDPKSNRIIR-GEKESDSTLSAITRLQD-BEENYKLIHDAVRLATEIIS 124  
 QY 115 LTGSDIHQIDITKNDPVGGILALSSHDTLKNVGDG-TTATIDRKHWRALTPOMFKY 172  
 DB 125 IAKDKVNAVD-----VAIPAVDTIVHVNNDQEIHKPKRAEYVQGPQAFKL 174  
 QY 173 GMLRDAIC-KTGNPAVTDASAL--ELIHEKPKIVBGPDKIKITRPEDLALA 223  
 DB 175 GTLAKAVDIYQGGIEGTCCSTVLKTLPEERVGIVSGSETNKLTPVDLFLA 228

RESULT 9  
 Q48154 PRELIMINARY; PRT; 474 AA.  
 ID Q48154  
 AC Q48154;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
 DE Bifunctional ribulose 5-phosphate reductase/CDP-ribitol  
 DE pyrophosphorylase.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9006;  
 RA Celis P., Moxon E., Byssen H., van Eldere J.;  
 RT "Genetic analysis of the Region II of the Haemophilus influenzae  
 RT serotype a capsulation locus."  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9006;  
 RA Celis P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9006;  
 RX MEDLINE=99194706; PubMed=10094675;  
 RA Follens A., Veiga-da-Cunha M., Merckx R., an Schaftingen E.,  
 RA van Eldere J.;  
 RT "Accl of Haemophilus influenzae type a Capsulation Locus Region II  
 RT encodes a bifunctional ribulose 5-phosphate reductase-CDP-ribitol  
 RT pyrophosphorylase."  
 RL J. Bacteriol. 181:2001-2007(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9006;  
 RA Follens A.;

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases  
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
(SDR) FAMILY.

DR EMBL: Z37516; CAAB5750.1; -  
DR InterPro: IPR002198; ADH short.  
DR InterPro: IPR001228; ISPD synthase.  
DR Pfam: PF00106; adh\_short\_1.  
DR Pfam: PF01128; ISPD\_1.  
DR PROSITE: PS00061; ADH\_SHORT; 1.  
KM OXidoreductase.  
SQ SEQUENCE 474 AA; 52438 MW; A8FBDF06F6452BF7 CRC64;

Query Match 21.5%; Score 258; DB 2; Length 474;  
Best Local Similarity 31.5%; Pred. No. 1,1e-13;

Matches 75; Conservative 36; Mismatches 85; Indels 42; Gaps 8;

DB 10 VPAAGVGRKMQADRPKQYLPPLAGKTVIEHTTLRLLESAPFQKVAVALSVEDPYMPELSI 69  
9 IILAGG:GSRMGIGYPRQFSKGIAGKTALEHTIPIQEHKEIDEIIVSE-----RTSY 61  
QY 70 AKRPDIITAR-----GKERADSVLSALKALEDIASENDVYLVDARP----- 113  
DB 62 RRIEDIVSVAGSGKRVRIIFGKERSDSTLSATLQD--EPRTKLIHDAVRPLATETI 120  
QY 114 ---CLTGSDIHLQIDTLKNDPVGGIILASGHDITLKVGDGT--ITATIDRKHWRALTFQ 168  
121 ISGCIAGKIDKRYAVD-----VAIRVDITIVVNDITGIIIMPKRAEYQGTQ 170  
QY 169 MEKYGMRLDMLQ-RTGKNPAVTDEASAL--ELIGHKPKIVEGRPNIKITRPEDIALA 223  
171 AFDGLGTLKAYADIIYTGIGIEGTDCSIVLKTPEERVGIVSGFETNIKITRPVDLPIA 228

## RESULT 10

Q8E7A0 PRELIMINARY; PRT; 238 AA.

ID Q8E7A0;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (EC 2.7.7.-).  
GN YGDP OR LA1048.  
OS Leptospiira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiroaceae; Leptospiira.  
OX NCBI\_TaxId=173;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;  
RA Ren S.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB011288; AAN48247.1; -  
KM Nucleosyltransferase; Transferase; Complete proteome.  
SQ SEQUENCE 238 AA; 26662 MW; C02400F042434DFB CRC64;

Query Match 21.4%; Score 256; DB 16; Length 238;  
Best Local Similarity 32.0%; Pred. No. 6,5e-14;

Matches 72; Conservative 48; Mismatches 91; Indels 14; Gaps 8;

QY 8 MAVVPAGVGRKMQADRPKQYLPPLAGKTVIEHTTLRLLESAPFQKVAVALSVEDPYMPELSI 67  
11 YVILAGG:GSRMGIGYPRQFSKGIAGKTALEHTIPIQEHKEIDEIIVSE-----RTSY 68  
QY 68 ---SLAKHPDIITAPGKERADSVLSALKALEDIASENDVYLVDARPCLTGSDIHL 122  
69 ESICASYLENEDEKRI-VQGGENRHSMLCGISVLD--FKBEDIILVHDAARPFVLADDELDS 125  
DB 123 QIDTLKNDPVGGIILASGHDITLKVGDGT--ITATIDRKHWRALTFQKVAVALSVEDPYMPELSI 181  
126 LCEKVASDGIATLASRTSETVLELNGKYV-SFLDEHVPWPKTPGCGIGDVLKELLTFS 184  
QY 182 TEGKNPAVTDEASALELIGHKPKIVEGRPNIKITRPEDIALA 226  
185 VDSIP--TDLCSWALTFGKTSIVESGNPLINKITRPEDIALAEVF 227

## RESULT 11

Q8G0H4 PRELIMINARY; PRT; 451 AA.

ID Q8G0H4;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase/4-diphosphocytidyl-2C-methyl-D-erythritol synthase.  
GN Brucella suis.  
OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxId=29461;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1330 / Biovar 1;  
RX MEDLINE=22247741; PubMed=12271122;  
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F., Read T.D., Dodson R.J., Umeyam L., Brinkac L.W., Beanan M.J., Tauberly S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R., Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E., Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L., Hoover D.L., Lindler L.B., Hailing S.M., Boyle S.M., Fraser C.M.;  
RT "The Brucella suis genome reveals fundamental similarities between animal and plant pathogens and symbionts";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).  
DR EMBL: AE014413; AAN30040.1; -  
KM Complete proteome.  
SQ SEQUENCE 451 AA; 48492 MW; 3BDB7285C056A161 CRC64;

Query Match 20.8%; Score 249.5; DB 16; Length 451;  
Best Local Similarity 33.2%; Pred. No. 5,7e-13;

Matches 77; Conservative 38; Mismatches 98; Indels 19; Gaps 9;

QY 9 AVVPAGVGRKMQADRPKQYLPPLAGKTVIEHTTLRLLESAPFQKVAVALSVEDPYMPELSI 66  
59 AVIVAGG:GSRMGIGYPRQFSKGIAGKTALEHTIPIQEHKEIDEIIVSE-----RTSY 118  
QY 67 LSI-AKHPDIITAPGKERADSVLSALKALEDIASENDVYLVDARPCLTGSDIHLQID 125  
119 RAVPEKHENVILVYTGPRGSTRUGILALXDEAPQ--YVLIHGVRFPI-GQDL-LERI 174  
DB 126 TLKNDPVGGIILASGHDITLKVGDGT--ITATIDRKHWRALTFQKVAVALSVEDPYMPELSI 177  
175 IANLTPDNGVLPALAVPDLTLRAADGAVETITISRTGFLPAOTPOAPFPAPPLIDHEKAF 234  
QY 178 ALORTEGNPAVTDEASALELIGHKPKIVEGRPNIKITRPEDIALA 229  
225 AINRD----FTDDALIAEWQEIIVRIRGSADNITKLFWADIEWADRLRQ 282

## RESULT 12

Q8RKI9 PRELIMINARY; PRT; 237 AA.

ID Q8RKI9;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Putative glycerol 5-phosphate cytidyltransferase.  
GN TARI.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxId=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M23;  
RA MEDLINE=21877339; PubMed=11882717;  
RA Lazarevic V., Abelian F.X., Beggs S., Karamata D., Manuel C.;  
RT "Comparison of ribitol and glycerol tetroic acid genes in Bacillus

RT subtilis W23 and 168: identical function, similar divergent  
RT organization, but different regulation."  
RL Microbiology 148:815-824(2002).  
DR EMBL: AJ313428; CAC86109.1; --  
DR InterPro: IPR001228; ISP1\_synthase.  
DR Pfam: PF01128; ISP1; 1.  
DR PROSITE: PS01295; ISP1; 1.  
KW transferase; Nucleoside diphosphate transferase.  
SQ SEQUENCE 237 AA; 26848 MW; 0B03F2F41DCE5AB CRC64;

Query Match 20.7%; Score 248.5; DB 2; Length 237;  
Best Local Similarity 28.7%; Pred. No. 2.8e-13;  
Matches 70; Conservative 47; Mismatches 94; Indels 33; Gaps 9;

QY 8 WAWPAGVGKRM-QADRPKQYPLAGKTVIEHTLRLLESDAFQKVAIVSVDPPYME 66  
DB 3 YAEILAGGKGRGNVNMKQPLKRPILITVTKPFLNRPDKLL--VSPKRWIN 59  
QY 67 LS-----IAKHEDITAPGKERADSVLSAKLEDI--ASENDWLVHDAAPCLTQ 117  
DB 60 HTKDLKKEFTIGDDRLVAVVGSGDRNESIMSGIRYEKFEFGIQDNDVITTHDSVPFLT- 118  
QY 118 SDHILQIDTLKNDPVGILLALSSHTLKKVDGTTATID-----RKGWMLTQ 168  
DB 119 ---HRIID---ENIDAVIQGVAVTVICAT-DTIIASEDOEFISDIPRDNMYQQTFO 170  
QY 169 MEKYGMELDALOR--TEGNPAVTDENASALELGHKPKIVEGRPNKIKITRPDLALQFY 226  
DB 171 SEPISTLVELYNNLSDQKAVLTLDAKICSLAGEKVLVRGEVFNKVTTPDLKAVANI 230  
QY 227 MEQO 230  
DB 231 LOER 234

## RESULT 13

08FM13

PRELIMINARY; PRT; 248 AA.

AC 08FM13;  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
DE Conserved hypothetical protein.  
GN CE2521.  
OS Corynebacterium efficiens.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Corynebacterium.  
OX NCBI\_TaxID=152794;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
RA Kawababayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,  
RA Usuda Y., Sugimoto S.;  
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF005222; BAC19331.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 248 AA; 26727 MW; AC66FF58948CEFF3 CRC64;

Query Match 20.7%; Score 248.5; DB 16; Length 248;  
Best Local Similarity 31.4%; Pred. No. 3e-13;  
Matches 72; Conservative 47; Mismatches 93; Indels 17; Gaps 7;

QY 9 AAVPAGVGKRM-QADRPKQYPLAGKTVIEHTLRLLESDAFQKVAIVSVDPPYME 67  
DB 8 ALTAAGGKGRGNVNMKQPLKRPILITVTKPFLNRPDKLL--VSPKRWIN 67  
QY 68 -----SIKAPD---ITAPGKERADSVLSAKLEDIASENDWLVHDAAPCLTGSDI 120  
DB 68 LRRRLGLNDPEGKRVLVHGGGRADSVAGLQALD--HDAIVLHDSAR--ALTPPGM 123  
QY 121 HQIDTLKNDPVGILLALSSHTLKKVDGTTATIDRKHVRLATPOMKYGKGLRAL 179

DB 124 IARVAVADGAPAVIPVPAVDTLKRVDSGVTDTPRADLRVQTFGLSLRAAN 183  
QY 180 QRTENP-----AVTDEASALELGHKPKIVEGRPNKIKITRPDLALQ 224  
DB 184 EKFAFPDGPFLPTDDASLMEWYGVETCVGGDPAFVPTTIDIMMLAO 232

## RESULT 14

08Y832

PRELIMINARY; PRT; 236 AA.

ID 08Y832  
AC 08Y832;  
DT 01-MAR-2002 (TREMblrel. 20, Created)  
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
DE 01-MAR-2002 (TREMblrel. 23, Last annotation update)  
DE Hypothetical protein Im01086.  
GN LMO1086.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EGD-e / Serovar 1/2a;  
RX MEDLINE=21537279; Pubmed=11679669;  
RA Glaeser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Darvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft U., Kuhn M., Kunst F., Kurpakt G.,  
RA Madano E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,  
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Putcell R.,  
RA Remmel B., Rose M., Schluteter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
RT "Comparative genomics of Listeria species.";  
RL Science 294:849-852(2001)  
DR EMBL; AL591977; CAC93164.1; --  
DR Listeria; LMO1086;  
DR InterPro: IPR001228; ISP1\_synthase.  
DR Pfam: PF01128; ISP1; 1.  
DR PROSITE: PS01295; ISP1; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 236 AA; 26712 MW; ACD3CE124E9A5098 CRC64;

Query Match 20.5%; Score 246; DB 16; Length 236;  
Best Local Similarity 29.5%; Pred. No. 4.6e-13;  
Matches 72; Conservative 44; Mismatches 94; Indels 34; Gaps 7;

QY 8 WAWPAGVGKRM-QADRPKQYPLAGKTVIEHTLRLLESDAFQKVAIVSVDPPYME 66  
DB 3 YAEILAGGKGRGNVNMKQPLKRPILITVTKPFLNRPDKLL--VSPKRWIN 62  
QY 67 LSIANK---PPIIAPGKERADSVLSAKLEDI--ASENDWLVHDAAPCLTGSDIH 121  
DB 63 DIKKYIDRSVIVIEGDTNRNTINGIRYKEFGNEDDIIVTDVAPFTTHRIE 122  
QY 122 LQIDTLKNDPVGILLALSSHT-----LKHVDGTTATIDRKHVRLATPOMK- 171  
DB 123 ENIM-----ALEGSDVTVIPAVDTIVESTNHPITDIPRGNVYQQTQOSRMK 174  
QY 172 -----YGLRDALORTGNPAVTDENASALELGHKPKIVEGRPNKIKITRPDLALQFY 226  
DB 175 TIQKHNNLTD-----DEKQILTDACKICLLAGKVLVNGGIGNIKITTPYDLKAVANI 229  
QY 227 MEQO 230  
DB 230 VOER 233

## RESULT 15

08KCU3

PRELIMINARY; PRT; 246 AA.

AC Q8KCU3;  
 DT 01-OCT-2002 (TRENBL-rel. 22, Created)  
 DT 01-OCT-2002 (TRENBL-rel. 22, Last sequence update)  
 DT 01-MAR-2003 (TRENBL-rel. 23, Last annotation update)  
 DE 4-diphosphocytidylyl-2C-methyl-D-erythritol synthase.  
 GN ISPD OR CT1317;  
 OS Chlorobium tepidum.  
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
 OC Chlorobium.  
 OX NCBI\_TaxId=1097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TLS / ATCC 49652 / DSM 12025;  
 RA MEDLINE=22103685; PubMed=12093901;  
 RA Eissen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
 RA Dodson R.J., Deboy R., Gwin M.L., Nelson W.C., Haft D.H.,  
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,  
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,  
 RA Vamathevan J., Knout H., White O., Gruber T.M., Ketchum K.A.,  
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.,  
 RA "The complete genome sequence of Chlorobium tepidum TLS, a  
 RA photosynthetic, anaerobic, green-sulfur bacterium.";  
 RA Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
 DR EMBL; AE012891; AAM72547.1; -.  
 DR TIGR; CT1317; -.  
 DR InterPro; IPR001228; ISPD\_synthase.  
 DR Pfam; PF01128; ISPD; 1.  
 DR PROSITE; PS01295; ISPD; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 246 AA; 27022 MW; D8DA46B55C2F5A56 CRC64;  
 Query Match 19.4%; Score 232.5; DB 16; Length 246;  
 Best Local Similarity 33.0%; Pred. No. 6.9e-12;  
 Matches 76; Conservative 39; Mismatches 98; Indels 17; Gaps 9;  
 QY 10 VVPAGVGGRMGAD--RPKQYPLAGKTYIEHTITRLLESDFQKVAVAISVED-PYWP 66  
 DB 6 IIAAGVGGRMLDGGRSKQMLEIGQPVYHTMKAFOEASTVESVYATLPDSIPVFE 65  
 QY 67 LSIANGPDITTA--PGKERADSVLSALXALEDIASEN---DWTLVHDAARPCVTGSDI 120  
 DB 66 IAKANGFTKITIIEGKERQDSIGNCKMLIEQEIENGSGVMPDALVHGAPFIQPEEI 125  
 QY 121 HQIDPTLKNDPVGGLIASSHTLKAV--DSDPTTATIDRKAVRALTPQWFKYGLRDA 178  
 DB 126 D-DIARLSATHGACVPATKPKDTIKYVGNPEIFGSTITDRSLRLQVOTPOGFAPAKLIEA 184  
 QY 179 LQRTGNPA--VTDEASALELGHKP--XIVGRPDNIKITRPEDLAAQ 224  
 DB 185 -HRLAGEQWYATDDALVVERYPQOALRIYETGYHNKITTPEVDVFIGE 233

Search completed: January 29, 2004, 15:54:06  
 Job time : 19.2185 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 : Search time 19.4547 Seconds  
(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947a-12

Perfect score: 1516  
Sequence: 1 MDVYAGGWRMPADAKLNTM.....LVPLAKLNGSALYKLEOG 285

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEW\_PUB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| 1          | 1516  | 100.0       | 285    | 10    | US-09-934-903-8     |
| 2          | 1516  | 100.0       | 285    | 10    | US-09-934-868-74    |
| 3          | 1516  | 100.0       | 285    | 11    | US-09-941-947a-12   |
| 4          | 638   | 42.1        | 279    | 12    | US-10-369-493-13766 |
| 5          | 620   | 40.6        | 279    | 12    | US-10-369-493-201   |
| 6          | 616   | 40.6        | 283    | 12    | US-10-369-493-802   |
| 7          | 584.5 | 38.5        | 280    | 12    | US-10-369-493-21079 |
| 8          | 584   | 38.5        | 294    | 12    | US-10-369-493-15324 |
| 9          | 582   | 38.4        | 276    | 12    | US-10-369-493-15690 |
| 10         | 582   | 38.4        | 276    | 12    | US-10-369-493-16082 |
| 11         | 580.5 | 38.3        | 298    | 12    | US-10-369-493-17751 |
| 12         | 572   | 37.7        | 270    | 12    | US-10-369-493-9184  |
| 13         | 572   | 37.7        | 270    | 12    | US-10-369-493-9359  |
| 14         | 556   | 36.7        | 213    | 12    | US-10-369-493-19675 |
| 15         | 533   | 35.2        | 233    | 12    | US-10-369-493-8419  |

| 16 | 407   | 26.8 | 284 | 12 | US-10-369-493-9956  | Sequence 9956, Ap |
|----|-------|------|-----|----|---------------------|-------------------|
| 17 | 341   | 22.5 | 289 | 12 | US-10-369-493-16541 | Sequence 16541, A |
| 18 | 327   | 21.6 | 269 | 12 | US-10-369-493-7851  | Sequence 7851, Ap |
| 19 | 324   | 21.4 | 289 | 12 | US-10-369-493-23046 | Sequence 23046, A |
| 20 | 320   | 21.1 | 257 | 12 | US-10-369-493-9660  | Sequence 9660, Ap |
| 21 | 317   | 20.9 | 287 | 12 | US-10-369-493-17097 | Sequence 17097, A |
| 22 | 292   | 19.3 | 316 | 12 | US-10-369-493-20265 | Sequence 20265, A |
| 23 | 286.5 | 18.9 | 295 | 12 | US-10-369-493-12283 | Sequence 12283, A |
| 24 | 281   | 18.5 | 297 | 15 | US-10-156-761-11123 | Sequence 11123, A |
| 25 | 278   | 18.3 | 311 | 12 | US-10-128-7131-8    | Sequence 8, Appl  |
| 26 | 274   | 18.1 | 292 | 12 | US-10-264-213-241   | Sequence 241, App |
| 27 | 273.5 | 18.0 | 261 | 12 | US-10-369-493-11764 | Sequence 11764, A |
| 28 | 273.5 | 18.0 | 261 | 12 | US-10-369-493-14419 | Sequence 14419, A |
| 29 | 273.5 | 18.0 | 261 | 12 | US-10-369-493-14670 | Sequence 14670, A |
| 30 | 273.5 | 18.0 | 261 | 12 | US-10-369-493-15150 | Sequence 15150, A |
| 31 | 270.5 | 17.8 | 315 | 12 | US-10-369-493-22607 | Sequence 2607, Ap |
| 32 | 267   | 17.6 | 305 | 12 | US-09-712-363-186   | Sequence 186, App |
| 33 | 254   | 16.8 | 240 | 12 | US-10-369-493-17829 | Sequence 17829, A |
| 34 | 250.5 | 16.5 | 268 | 12 | US-10-369-493-65    | Sequence 65, Appl |
| 35 | 239   | 15.8 | 295 | 12 | US-10-369-493-16874 | Sequence 16874, A |
| 36 | 233.5 | 15.4 | 271 | 12 | US-10-369-493-3024  | Sequence 3024, Ap |
| 37 | 216   | 14.2 | 311 | 10 | US-09-738-626-4515  | Sequence 4515, Ap |
| 38 | 203   | 13.4 | 245 | 12 | US-10-369-493-10645 | Sequence 10645, A |
| 39 | 178   | 11.7 | 141 | 12 | US-10-289-762-1236  | Sequence 1236, Ap |
| 40 | 160   | 10.6 | 99  | 12 | US-10-264-213-208   | Sequence 208, App |
| 41 | 127   | 8.4  | 238 | 12 | US-09-882-227-358   | Sequence 358, App |
| 42 | 125   | 8.2  | 314 | 9  | US-09-815-242-10963 | Sequence 10963, A |
| 43 | 116   | 7.7  | 575 | 12 | US-10-104-047-2815  | Sequence 2815, Ap |
| 44 | 116   | 7.7  | 813 | 12 | US-10-108-2604-3901 | Sequence 3901, Ap |
| 45 | 116   | 7.7  | 990 | 12 | US-10-094-749-2568  | Sequence 2568, Ap |

## ALIGNMENTS

RESULT 1  
US-09-934-903-8  
Sequence 8, Application US/09934903  
Patent No. US20020102690A1  
GENERAL INFORMATION:  
APPLICANT: Koffas, Mattheos  
APPLICANT: Odum, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. US20020102690A1, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Rouviere, Pierre  
APPLICANT: Piccataggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
FILE REFERENCE: C1646 US NA  
CURRENT APPLICATION NUMBER: US/09/934, 903  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229, 907  
PRIOR FILING DATE: September 1, 2001  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 8  
LENGTH: 285  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ORF4  
US-09-934-903-8

Query Match 100.0%; Score 1516; DB 10; Length 285;  
Best Local Similarity 100.0%; Pred. No. 5.7e-155;  
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDVYAGGWRMPADAKLNTMRTGRRPDGYHLOTFQMDLCMDLTFEPVDDGVTAR 60  
DB 1 MDVYAGGWRMPADAKLNTMRTGRRPDGYHLOTFQMDLCMDLTFEPVDDGVTAR 60  
QY 61 NPISVPEQDDLTFRANLTKSHTCVRCVCIDIERKLPMSGGGLGGSSDAATTVLVNR 120

```

Db      61 NPISGVPEDDDLTVRAANLKSHTCGVGCIDIEKRLPMGGGLGGSSDAATLVNLR 120
Qy      121 LMGJLSKREIMDLGLRLGADVVFVFGCSAMGEGVSEDLOATITLPEQMFVILKPDCHVN 180
Db      121 LMGJLSKREIMDLGLRLGADVVFVFGCSAMGEGVSEDLOATITLPEQMFVILKPDCHVN 180
Qy      181 TGEIFSAENLIRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVKDAIDALLCYAARLGT 240
Db      181 TGEIFSAENLIRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVKDAIDALLCYAARLGT 240
Qy      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEQG 285
Db      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEQG 285

RESULT 2
US-09-934-868-74
; Sequence 74, Application US/09934868
; Patent No. US20020137190A1
; GENERAL INFORMATION:
; APPLICANT: Koffas, Mattheos
; APPLICANT: Odom, James M.
; APPLICANT: Schenck, Andreas J.
; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN
; FILE REFERENCE: C1A596 US NA
; CURRENT APPLICATION NUMBER: US/09/934,868
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/229,858
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 74
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Methylobionas 16a
; FEATURE:
; OTHER INFORMATION: Amino acid sequences encoded by ISPE
US-09-934-868-74

Query Match      100.0%; Score 1516; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. No. 5.7e-156;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MDYAAGWGERMPAPAKNIMLRITGRRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
Db      1 MDYAAGWGERMPAPAKNIMLRITGRRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
Qy      61 NPISGVPEDDDLTVRAANLKSHTCGVGCIDIEKRLPMGGGLGGSSDAATLVNLR 120
Db      61 NPISGVPEDDDLTVRAANLKSHTCGVGCIDIEKRLPMGGGLGGSSDAATLVNLR 120
Qy      121 LMGJLSKREIMDLGLRLGADVVFVFGCSAMGEGVSEDLOATITLPEQMFVILKPDCHVN 180
Db      121 LMGJLSKREIMDLGLRLGADVVFVFGCSAMGEGVSEDLOATITLPEQMFVILKPDCHVN 180
Qy      181 TGEIFSAENLIRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVKDAIDALLCYAARLGT 240
Db      181 TGEIFSAENLIRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVKDAIDALLCYAARLGT 240
Qy      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEQG 285
Db      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEQG 285

RESULT 3
US-09-941-947A-12
; Sequence 12, Application US/09941947A
; Publication No. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: Dicosimo, Deana J.

```

```

; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odom, J. Martin
; APPLICANT: Picataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: C1A903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Methylobionas 16a
US-09-941-947A-12

Query Match      100.0%; Score 1516; DB 11; Length 285;
Best Local Similarity 100.0%; Pred. No. 5.7e-156;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MDYAAGWGERMPAPAKNIMLRITGRRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
Db      1 MDYAAGWGERMPAPAKNIMLRITGRRPDGYHLQTVFQMLDCDMLTFHPVDGRTLR 60
Qy      61 NPISGVPEDDDLTVRAANLKSHTCGVGCIDIEKRLPMGGGLGGSSDAATLVNLR 120
Db      61 NPISGVPEDDDLTVRAANLKSHTCGVGCIDIEKRLPMGGGLGGSSDAATLVNLR 120
Qy      121 LMGJLSKREIMDLGLRLGADVVFVFGCSAMGEGVSEDLOATITLPEQMFVILKPDCHVN 180
Db      121 LMGJLSKREIMDLGLRLGADVVFVFGCSAMGEGVSEDLOATITLPEQMFVILKPDCHVN 180
Qy      181 TGEIFSAENLIRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVKDAIDALLCYAARLGT 240
Db      181 TGEIFSAENLIRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVKDAIDALLCYAARLGT 240
Qy      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEQG 285
Db      241 GACVFAQFCNKEDAESALEGLKDRMLVFLAKGLNOSALYKLEQG 285

RESULT 4
US-10-369-493-13766
; Sequence 13766, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10 (52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13766
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13766

Query Match      42.1%; Score 638; DB 12; Length 279;
Best Local Similarity 49.6%; Pred. No. 1.2e-60;
Matches 135; Conservative 41; Mismatches 96; Indels 0; Gaps 0;

```

QY 12 PAKPKNTLMIRIGRRPDEGTHLLQTFQFMOLDCMTLTFHPVDGRVTLRNPLSGVPEODD 71  
Db 5 PSPAKNTLMHLIGRRDGTHEHQLTFQFLPDYDEDTTFAYRVDGYIRLHTFEDGYPHOSH 64  
QY 72 LTYRANMLKSHTGCVRGVCIDIERNLPMGGAGGGSSDAATTVLVNLMGLGSKREL 131  
Db 65 LTYRAAALQOBQASCSIGIDIWIDIKLTPMGGGGGSSNAATTLGLNHLWRLGMDBDRL 124  
QY 132 MDYGLRIGADVPVFFVFGCSAMGEVGESEDIQATLTPPQFVILIKPQCCHNTGTFPSAEMLT 199  
Db 125 AAGLTIGADVPVFFVFGHAAFAAGVGEKLTLPBPAPFWVVLVPQVSFTARISDFLLT 188  
QY 192 RNSAVVTMSDFLAGDNRENCSEVYCKLYRPVKDAIDALICVAEARLTGTGA CVFAQFCNK 251  
Db 185 RNPPLIKVREVPBGNRSNDCLPVASRRYREYVRNALNLGNFTEAKLTGTGSCVFGGFPSSK 244  
QY 252 EDASALEGKXDRWLVFLAKGLNQSLYYKKLE 263  
Db 245 AEAADKVSALLTETLTGTGVAKGSSVSLTHKKLQ 276

```

RESULT 5
US-10-369-493-201
: Sequence 201, Application US/10369493
: Publication No.: US20030233675A1
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
: PII# REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: CURRENT FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 201
: LENGTH: 279
: TYPE: PRT
: ORGANISM: Xenorhabdus nematophilus
: US-10-369-493-201

```

|    | Query Match           | 40.9%;   | Score 620;          | DB 58;    | length 279; |
|----|-----------------------|--|---------------------|-----------|-------------|
|    | Best Local Similarity | 48.0%;   | Pred. No. 1; Le 12; |           |             |
|    | Matches 132;          | Conservative 36;   | Mismatches 101;     | Indels 6; | Gaps 2      |
| QY | 11                    | WPAAPALNTLMTRTGRPDPTHTLLCTVQMTLDCWLTFFHVEYDGRVTLANPISGVPEOD    | 70                  |           |             |
| Db | 5                     | WSPAPALNTFLVITGTGRSDGYHRLQTLRFQFLDYGEIITTRQDNQIRLLPLPVKGVND    | 64                  |           |             |
| QY | 71                    | DLTVRAANTLKSH-----TGC--VRQVCIDIEKNLPMGGGLGGSSDAATTVVLRNLWGL    | 124                 |           |             |
| Db | 65                    | NLIVRAARLLQDHAHQSGTQYKRADHIHNRLEPMGGSLGGSSNAATVILALNHYMOT      | 124                 |           |             |
| QY | 125                   | GLSKRELMDLGLRLGADVVFVFGCSAMGEGVEDLQAITLPEQMPVITIKPDGVNTGRT     | 184                 |           |             |
| Db | 125                   | NMSDDSLAQGLVTLGADVVFVFKGMAAAREGIEGLQSPASBEKFWFLVHNGIIEIPPKI    | 184                 |           |             |
| QY | 185                   | FSALNLTNRNSAVVWMSDFLAGDNNDCEBVCIKLRPVKDAIDALLCAEABRLTGTGACV    | 244                 |           |             |
| Db | 185                   | FTDPBELKRNSSPIRITLPAALLQAPFNKDCPIARKRRREVEGLSMLLBETTSRLTGTGACV | 244                 |           |             |
| QY | 245                   | FAQFCNKEDASALEGLDNRMLVFLAKGLNSALT                              | 279                 |           |             |
| Db | 245                   | FGFESBPASARKVTLQAPFMQGFVARKVNSPSLH                             | 279                 |           |             |

RESULT 6  
US-10-369-493-802

; Sequence 802, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ. ID NOS: 47374  
 ; SEQ. ID NO 802  
 ; LENGTH: 283  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; US-10-369-493-802

[illegible]

```

RESULT 7
US-10-369-493-21079
; Sequence 21079, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ. ID NO 21079
; LENGTH: 280
; TYPE: prt
; ORGANISM: Xenorhabdus nematophilus
; US-10-369-493-21079

```

|                       |        |                    |        |             |
|-----------------------|--------|--------------------|--------|-------------|
| Query Match           | 38.6%; | Score 584.5;       | DB 12; | Length 280; |
| Best Local Similarity | 45.3%; | Pred. No. 7.8e-55; |        |             |



Matches 125; Conservative 42; Mismatches 102; Indels 7; Gaps 2;

QY 11 WPAPAKLMLRTGRRPDGTHLLQTVFQMLDLCMTLTFHVPDDGRV-TLRNPISGVPEQ 70  
 DB 5 WSPAPAKLMLRTGRRPDGTHLLQTVFQMLDLCMTLTFHVPDDGRV-TLRNPISGVPEQ 64  
 QY 71 DLTVRAANLTKSH-----TGCVR-GVCIDIEKNLPMGGGLGGSSDAATTVLNLWGLSKR 123  
 DB 65 NLIVRAARLQDHLKONNKGTHGLADTHIDKLPNGGGLGGSSDAATTVLNLWGLSKR 124  
 QY 124 LGSIKRLMDLGRIGADVPPVFGCSAMGEGVEDLQATLPEQWFIKKPDCHVNTGE 183  
 DB 125 ANLSDEDLQIGISIGADVPPVFGCSAMGEGVEDLQATLPEQWFIKKPDCHVNTGE 184  
 QY 184 LFSKNTLRNSAVTMSDFLAGNRNDCSEVCKLYRPVKDAI-DALLCYEARLTGTGACVPAQF 243  
 DB 185 LFTDPELKRNSPIKSLPALLQAPFTNDCEIARRRFRKVKQLLSWLEVPASRLTGTGAC 244  
 QY 244 VFAQCNKEDAESALBGLKDRMLVFLAKGLNOSALY 279  
 DB 245 VFGFSEASARAKLNLNOSPMMOGFVARGVNIISPLH 280

RESULT 8  
 US-10-369-493-15324  
 : Sequence 15324, Application US/10369493  
 : Publication No. US20030233675A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cao, Yongwei  
 : APPLICANT: Hinkle, Gregory J.  
 : APPLICANT: Slater, Steven C.  
 : APPLICANT: Goldman, Barry S.  
 : APPLICANT: Chen, Xianfeng  
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 : FILE REFERENCE: 38-10(52052)B  
 : CURRENT APPLICATION NUMBER: US/10/369,493  
 : PRIOR FILING DATE: 2003-02-28  
 : PRIOR APPLICATION NUMBER: US 60/360,039  
 : PRIOR FILING DATE: 2002-02-21  
 : NUMBER OF SEQ ID NOS: 47374  
 : SEQ ID NO 15324  
 : LENGTH: 294  
 : TYPE: PRT  
 : ORGANISM: Xanthomonas campestris  
 : US-10-369-493-15324

Query Match 38.5%; Score 584; DB 12; Length 294;  
 Best Local Similarity 45.8%; Pred. No. 9.5e-55;  
 Matches 126; Conservative 42; Mismatches 105; Indels 2; Gaps 2;

QY 11 WPAPAKLMLRTGRRPDGTHLLQTVFQMLDLCMTLTFHVPDDGRV-TLRNPISGVPEQ 69  
 DB 17 WSPAPAKLMLRTGRRPDGTHLLQTVFQMLDLCMTLTFHVPDDGRV-TLRNPISGVPEQ 76  
 QY 70 DLTVRAANLTKSHGTCVGVGVCIDIEKNLPMGGGLGGSSDAATTVLNLWGLSKR 123  
 DB 77 DDLWRAARALQIHAGTALGAEIRYDKRIPAGGCGGSSDAATTVLNLWGLSKR 136  
 QY 130 ELMDLGRIGADVPPVFGCSAMGEGVEDLQATLPEQWFIKKPDCHVNTGEISAE 189  
 DB 137 TLAEIGLRIGADVPPVFGCSAMGEGVEDLQATLPEQWFIKKPDCHVNTGEISAE 196  
 QY 190 LTRNSAVTMSDFLAGNRNDCSEVCKLYRPVKDAI-DALLCYEARLTGTGACVPAQF 248  
 DB 197 LTRDAAPAKIADPAGSSLLDNAPFVLRREPAIEAVFQALSRIGTRLTGSGSCFVEF 256  
 QY 249 CNKEDAESALBGLKDRMLVFLAKGLNOSALY 283  
 DB 257 ATRAAEQAMAHLPENLRAMWVEGAHSPLLDALD 291

RESULT 9

US-10-369-493-15690  
 : Sequence 15690, Application US/10369493  
 : Publication No. US20030233675A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cao, Yongwei  
 : APPLICANT: Hinkle, Gregory J.  
 : APPLICANT: Slater, Steven C.  
 : APPLICANT: Goldman, Barry S.  
 : APPLICANT: Chen, Xianfeng  
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 : FILE REFERENCE: 38-10(52052)B  
 : CURRENT APPLICATION NUMBER: US/10/369,493  
 : PRIOR FILING DATE: 2003-02-28  
 : PRIOR APPLICATION NUMBER: US 60/360,039  
 : PRIOR FILING DATE: 2002-02-21  
 : NUMBER OF SEQ ID NOS: 47374  
 : SEQ ID NO 15690  
 : LENGTH: 276  
 : TYPE: PRT  
 : ORGANISM: Xanthomonas campestris  
 : US-10-369-493-15690

Query Match 38.4%; Score 582; DB 12; Length 276;  
 Best Local Similarity 46.0%; Pred. No. 1.4e-54;  
 Matches 126; Conservative 41; Mismatches 105; Indels 2; Gaps 2;

QY 11 WPAPAKLMLRTGRRPDGTHLLQTVFQMLDLCMTLTFHVPDDGRV-TLRNPISGVPEQ 69  
 DB 3 WSPAPAKLMLRTGRRPDGTHLLQTVFQMLDLCMTLTFHVPDDGRV-TLRNPISGVPEQ 62  
 QY 70 DLTVRAANLTKSHGTCVGVGVCIDIEKNLPMGGGLGGSSDAATTVLNLWGLSKR 129  
 DB 63 DDLWRAARALQIHAGTALGAEIRYDKRIPAGGCGGSSDAATTVLNLWGLSKR 122  
 QY 130 ELMDLGRIGADVPPVFGCSAMGEGVEDLQATLPEQWFIKKPDCHVNTGEISAE 189  
 DB 123 TLAEIGLRIGADVPPVFGCSAMGEGVEDLQATLPEQWFIKKPDCHVNTGEISAE 182  
 QY 190 LTRNSAVTMSDFLAGNRNDCSEVCKLYRPVKDAI-DALLCYEARLTGTGACVPAQF 248  
 DB 183 LTRDAAPAKIADPAGSSLLDNAPFVLRREPAIEAVFQALSRIGTRLTGSGSCFVEF 242  
 QY 249 CNKEDAESALBGLKDRMLVFLAKGLNOSALY 282  
 DB 243 ATRAAEQAMAHLPENLRAMWVEGAHSPLLDALD 276

RESULT 10  
 US-10-369-493-16082  
 : Sequence 16082, Application US/10369493  
 : Publication No. US20030233675A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cao, Yongwei  
 : APPLICANT: Hinkle, Gregory J.  
 : APPLICANT: Slater, Steven C.  
 : APPLICANT: Goldman, Barry S.  
 : APPLICANT: Chen, Xianfeng  
 : TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 : FILE REFERENCE: 38-10(52052)B  
 : CURRENT APPLICATION NUMBER: US/10/369,493  
 : PRIOR FILING DATE: 2003-02-28  
 : PRIOR APPLICATION NUMBER: US 60/360,039  
 : PRIOR FILING DATE: 2002-02-21  
 : NUMBER OF SEQ ID NOS: 47374  
 : SEQ ID NO 16082  
 : LENGTH: 276  
 : TYPE: PRT  
 : ORGANISM: Xanthomonas campestris  
 : US-10-369-493-16082

Query Match 38.4%; Score 582; DB 12; Length 276;

Best Local Similarity 46.0%; Pred. No. 1.4e-54;  
Matches 126; Conservative 41; Mismatches 105; Indels 2; Gaps 2;

QY 11 WPAPAKLIMLRITGRPDGYHLQTVFQMLDCLMTLFFPVDDGRV-TLRNPISGVPEQ 69  
Db 3 WPAPAKLIMLRITGRPDGYHLQTVFQMLDCLMTLFFPVDDGRV-TLRNPISGVPEQ 62  
QY 70 DDLTVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSKR 129  
Db 63 DDLTVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSKR 122  
QY 130 ELMDLGRLGADVPVFGSGMWGSEVSEDQATLTPQWVILIKEDCVNTEIFSAEN 189  
Db 123 TLAEGLRLGADVPVFGSGMWGSEVSEDQATLTPQWVILIKEDCVNTEIFSAEN 182  
QY 190 LTRNSAVVTMSDFLAGDNNDSCSEVCKLYRPYKDAI-DALLCYAEARLTGTGACVPAQF 248  
Db 183 LTRNSAVVTMSDFLAGDNNDSCSEVCKLYRPYKDAI-DALLCYAEARLTGTGACVPAQF 242  
QY 249 CNKEDASALEGLKDRMLVFLAKGLNQSALYKKL 282  
Db 243 ATRAAEQAMHLPGMLRAVVEGAHSPLDAL 276

## RESULT 11

US-10-369-493-17751  
Sequence 17751, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 17751  
LENGTH: 298  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-17751

Query Match 38.3%; Score 580.5; DB 12; Length 298;  
Best Local Similarity 46.3%; Pred. No. 2.3e-54;  
Matches 132; Conservative 44; Mismatches 106; Indels 3; Gaps 3;

QY 1 MDYAAWGERPAPAKLIMLRITGRPDGYHLQTVFQMLDCLMTLFFPVDDGRV-TL 59  
Db 11 VDDGVGM-SAMPAPAKLIMLRITGRPDGYHLQTVFQMLDCLMTLFFPVDDGRV-TL 69  
QY 60 RNPISGVPEQDILTVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNL 119  
Db 70 GESVTVVEADLVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNL 129  
QY 120 RLMGLSKRLEIMDLGRLGADVPVFGSGMWGSEVSEDQATLTPQWVILIKEDCVN 179  
Db 130 ALMHLRLDVAVALGRLGADVPVFGSGMWGSEVSEDQATLTPQWVILIKEDCVN 189  
QY 180 NTGERFSANLTVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNL 238  
Db 190 PFRLEPDDILTVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNL 249  
QY 239 GTGACVPAQFCNKEDASALEGLKDRMLVFLAKGLNQSALYKKL 283  
Db 250 GSGSGCFVEFSTRDASALEGLKDRMLVFLAKGLNQSALYKKL 294

## RESULT 12

US-10-369-493-9184  
Sequence 9184, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 9184  
LENGTH: 270  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-9184

Query Match 37.7%; Score 572; DB 12; Length 270;  
Best Local Similarity 47.4%; Pred. No. 1.7e-53;  
Matches 128; Conservative 42; Mismatches 98; Indels 2; Gaps 2;

QY 11 WPAPAKLIMLRITGRPDGYHLQTVFQMLDCLMTLFFPVDDGRV-TLRNPISGVPEQ 69  
Db 1 WPAPAKLIMLRITGRPDGYHLQTVFQMLDCLMTLFFPVDDGRV-TLRNPISGVPEQ 60  
QY 70 DDLTVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSKR 129  
Db 61 DDLTVRAANLKSHTGCVGVICIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSKR 120  
QY 130 ELMDLGRLGADVPVFGSGMWGSEVSEDQATLTPQWVILIKEDCVNTEIFSAEN 189  
Db 121 TLAEGLRLGADVPVFGSGMWGSEVSEDQATLTPQWVILIKEDCVNTEIFSAEN 180  
QY 190 LTRNSAVVTMSDFLAGDNNDSCSEVCKLYRPYKDAI-DALLCYAEARLTGTGACVPAQF 248  
Db 181 LTRNSAVVTMSDFLAGDNNDSCSEVCKLYRPYKDAI-DALLCYAEARLTGTGACVPAQF 240  
QY 249 CNKEDASALEGLKDRMLVFLAKGLNQSALYKKL 283  
Db 241 STRDASALEGLKDRMLVFLAKGLNQSALYKKL 270

## RESULT 13

US-10-369-493-9359  
Sequence 9359, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 9359  
LENGTH: 270  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-9359

Query Match 37.7%; Score 572; DB 12; Length 270;  
 Best Local Similarity 47.4%; Pred. No. 1.7e-53;  
 Matches 128; Conservative 42; Mismatches 98; Indels 2; Gaps 2;

QY 11 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRV-TLRNPISGVEEQ 69  
 DB 1 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRV-TLRNPISGVEEQ 60  
 QY 70 DDLTVRANILKSHGCVRGVCIDIBKXNPMWGGGSSDAATLVVLANRIMGLSKR 129  
 DB 61 DDLTVRANILKSHGCVRGVCIDIBKXNPMWGGGSSDAATLVVLANRIMGLSKR 120  
 QY 130 BLMDGLRLGADVPPVFGCSAMGEGVEDLQATITLPEOMFVILKPDCHVNTGELFSAEN 189  
 DB 121 VLAALGLRLGADVPPVFGCSAMGEGVEDLQATITLPEOMFVILKPDCHVNTGELFSAEN 180  
 QY 190 LTRNSAVVTMSDFLAGDNNDCE-VVCKLTPVYDAIDALICVBAALITGTGACVPAOF 248  
 DB 181 LTRNSAVVTMSDFLAGDNNDCE-VVCKLTPVYDAIDALICVBAALITGTGACVPAOF 240  
 QY 249 CNKEDAESALBGLKXDRWLVFLAKGNOSAL 278  
 DB 241 STRDERECLERLPYGLCAWVADGASRSP 270

# RESULT 14

US-10-369-493-19675  
 ; Sequence 19675, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 19675  
 ; LENGTH: 213  
 ; TYPE: PRT  
 ; ORGANISM: Nitrosomonas europaea  
 US-10-369-493-19675

Query Match 36.7%; Score 556; DB 12; Length 213;  
 Best Local Similarity 51.6%; Pred. No. 6.6e-52;  
 Matches 110; Conservative 34; Mismatches 67; Indels 2; Gaps 2;

QY 11 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRVTLRNPISGVEEQ 70  
 DB 1 WPAPAKNIMLRITGRPRDGYHLLQTVFQMDLCMLTFHPVDDGRVTLRNPISGVEEQ 60  
 QY 71 DILVPAANILKSHGCVRGVCIDIBKXNPMWGGGSSDAATLVVLANRIMGLSKR 129  
 DB 61 DILVPAANILKSHGCVRGVCIDIBKXNPMWGGGSSDAATLVVLANRIMGLSKR 120  
 QY 130 BLMDGLRLGADVPPVFGCSAMGEGVEDLQATITLPEOMFVILKPDCHVNTGELFSAEN 189  
 DB 121 VLAALGLRLGADVPPVFGCSAMGEGVEDLQATITLPEOMFVILKPDCHVNTGELFSAEN 180  
 QY 190 LTRNSAVVTMSDFLAGDNNDCE-VVCKLTPVYDAIDALICVBAALITGTGACVPAOF 248  
 DB 181 LTRNSAVVTMSDFLAGDNNDCE-VVCKLTPVYDAIDALICVBAALITGTGACVPAOF 240

RESULT 15  
 US-10-369-493-8419  
 ; Sequence 8419, Application US/10369493

Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 8419  
 ; LENGTH: 233  
 ; TYPE: PRT  
 ; ORGANISM: Ralstonia metallidurans  
 US-10-369-493-8419

Query Match 35.2%; Score 533; DB 12; Length 233;  
 Best Local Similarity 49.1%; Pred. No. 2.4e-49;  
 Matches 114; Conservative 35; Mismatches 77; Indels 6; Gaps 2;

QY 53 DGRVTLRNPISGVEEQDILTVRANILKSHGCVRGVCIDIBKXNPMWGGGSSDA 112  
 DB 2 DGRVTLRNPISGVEEQDILTVRANILKSHGCVRGVCIDIBKXNPMWGGGSSDA 61  
 QY 113 TTVVLANRIMGLSKRDLGLRLGADVPPVFGCSAMGEGVEDLQATITLPEOMFV 172  
 DB 62 TTVVLANRIMGLSKRDLGLRLGADVPPVFGCSAMGEGVEDLQATITLPEOMFV 121  
 QY 173 IKPDCHVNTGELFSAENLITNSAVVTMSDFLAGDN-----RNDCEVVCCLYRPAKDAID 227  
 DB 122 IKPDCHVNTGELFSAENLITNSAVVTMSDFLAGDN-----RNDCEVVCCLYRPAKDAID 181  
 QY 228 ALICVBAALITGTGACVPAOFCKNEDAESALBGLKXDRWLVFLAKGNOSAL 278  
 DB 182 ALICVBAALITGTGACVPAOFCKNEDAESALBGLKXDRWLVFLAKGNOSAL 233

Search completed: January 29, 2004, 16:21:17  
 Job time : 20.4547 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 ; Search time 26.0514 Seconds  
(without alignments)  
1736.455 Million cell updates/sec

Title: US-09-941-947A-12

Sequence: 1 MDYAGWGERMPAPKLNLM.....LWFLAKGINSALYKKEEG 285

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_19jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 1516  | 100.0       | 285    | 23 | ABG61587    |
| 2          | 1516  | 100.0       | 285    | 23 | AAE22304    |
| 3          | 1516  | 100.0       | 285    | 23 | AAU80328    |
| 4          | 624   | 41.2        | 281    | 24 | ABP78450    |
| 5          | 616   | 40.6        | 283    | 21 | AAAB1374    |
| 6          | 616   | 40.6        | 283    | 22 | AAAB6286    |
| 7          | 332   | 21.9        | 283    | 23 | ABP25586    |
| 8          | 320   | 21.1        | 293    | 23 | ABBS0111    |
| 9          | 302   | 19.9        | 299    | 23 | ABP38503    |

|    |       |      |     |    |          |
|----|-------|------|-----|----|----------|
| 10 | 278   | 18.3 | 311 | 24 | AAE31685 |
| 11 | 277   | 18.3 | 401 | 22 | AAE68287 |
| 12 | 275   | 18.1 | 383 | 22 | AAE68288 |
| 13 | 270.5 | 17.8 | 315 | 22 | AAE68291 |
| 14 | 268   | 17.7 | 316 | 22 | AAE70486 |
| 15 | 268   | 17.7 | 383 | 22 | AAE70487 |
| 16 | 267   | 17.6 | 306 | 22 | AAE81135 |
| 17 | 263   | 17.3 | 405 | 22 | AAE68285 |
| 18 | 224   | 14.8 | 360 | 22 | ABG17595 |
| 19 | 223   | 14.7 | 324 | 22 | AAU54372 |
| 20 | 216   | 14.2 | 311 | 22 | AAE90761 |
| 21 | 216   | 14.2 | 311 | 22 | AAE76645 |
| 22 | 204.5 | 13.5 | 316 | 23 | ABP65646 |
| 23 | 195   | 12.9 | 184 | 21 | AAE12035 |
| 24 | 195   | 12.9 | 185 | 21 | AAE12034 |
| 25 | 178   | 11.7 | 141 | 20 | AAE35818 |
| 26 | 176.5 | 11.6 | 263 | 20 | AAE37333 |
| 27 | 140   | 9.2  | 746 | 22 | ABE29149 |
| 28 | 127   | 8.4  | 238 | 19 | AAE98746 |
| 29 | 125   | 8.2  | 314 | 22 | AAU53370 |
| 30 | 115.5 | 7.6  | 483 | 20 | AAE1016  |
| 31 | 114   | 7.5  | 244 | 21 | AAE45153 |
| 32 | 114   | 7.5  | 246 | 21 | AAE45152 |
| 33 | 114   | 7.5  | 370 | 21 | AAE45151 |
| 34 | 111.5 | 7.4  | 288 | 23 | ABP28086 |
| 35 | 110   | 7.3  | 229 | 21 | AAE16687 |
| 36 | 110   | 7.3  | 231 | 21 | AAE16686 |
| 37 | 110   | 7.3  | 355 | 21 | AAE16685 |
| 38 | 105   | 6.9  | 309 | 22 | AAU38118 |
| 39 | 101   | 6.7  | 310 | 22 | AAU34420 |
| 40 | 98    | 6.5  | 331 | 22 | AAE60866 |
| 41 | 98    | 6.5  | 341 | 22 | AAU53064 |
| 42 | 96    | 6.3  | 289 | 21 | AAE1774  |
| 43 | 96    | 6.3  | 289 | 22 | AAU37651 |
| 44 | 96    | 6.3  | 289 | 22 | AAU37989 |
| 45 | 96    | 6.3  | 289 | 24 | ABU01802 |

#### ALIGNMENTS

RESULT 1  
ID ABG61587  
ID ABG61587 standard; Protein, 285 AA.

AC ABG61587;  
XX  
DT 27-AUG-2002 (first entry)  
XX

DE High growth methanotrophic bacterial strain polypeptide #37.

XX High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme;  
KW nehanne; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;  
KW pyrophosphate dependent phosphofructokinase; nitrogen-containing compound;  
KW ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;  
KW methane-containing environment; waste water treatment system; isoprenoid;  
KW nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.

OS Methylobionas 16a.

PN W0200220728-A2.

PD 14-MAR-2002.

PF 28-AUG-2001; 2001WO-US26827.

PR 01-SEP-2000; 2000US-229858P.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

PI Koffas M, Odum JM, Schenzle A;

XX WPI; 2002-452200/48.

Rhodococcus erythr  
Amino acid sequenc  
Amino acid sequenc  
Amino acid sequenc  
A. thaliana YCNB p  
Arabidopsis thalia  
Mycobacterium tube  
Amino acid sequenc  
Novel human diagno  
Propionibacterium  
C. glutamicum prote  
Corynebacterium g1  
Bifidobacterium 10  
Arabidopsis thalia  
Arabidopsis thalia  
Amino acid sequenc  
Novel human diagno  
Novel human diagno  
H. pylori GPO 100  
Haemophilus infu  
Fragment of homose  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Streptococcus poly  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Salmonella typhi c  
E. coli cellular p  
Neuronal pathway  
Enterococcus faeca  
Streptococcus pneu  
Streptococcus pneu  
Streptococcus pneu  
Streptococcus pneu  
S. pneumoniae type

DR N-PSDB; ABRK3266.

XX New high growth methanotrophic bacterial strain, useful for producing  
PT single cell proteins, grows on a C1 carbon substrate, and comprises a  
PT functional gene encoding in Embden-Meyerhof carbon pathway

XX Claim 11; Page 149-150; 157pp; English.

XX The invention relates to a high growth methanotrophic bacterial strain,  
CC which grows on a C1 carbon substrate e.g. methane and methanol, and  
CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a  
CC gene coding a pyrophosphate dependent phosphotransferase enzyme or a 16S  
CC rRNA. The bacterial strain is useful for the production of single cell  
CC protein and for the biotransformation of a nitrogen-containing compound,  
CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the  
CC production of a feed product comprising a protein, carbohydrates and a  
CC pigment and for reducing oxygen demand, for removing nitrates and  
CC nitrites in methane-containing environments such as landfills, waste  
CC water treatment systems or anywhere that methane, oxygen and nitrates are  
CC present. The bacterial strain of the invention can be used as a  
CC denitrifying agent for the conversion of nitrate or nitrite to nitrous  
CC oxide with methane or methanol as a carbon source. It is also used in the  
CC production of biomass including proteins, carbohydrates and a wide  
CC variety of pigments (particularly for isoprenoid pigments for the  
CC purpose of generating animal feeds), in production of terpenoid and  
CC carotenoid compounds, useful as pigments and as monomers in polymeric  
CC materials and in production of exopolysaccharides at high levels.  
CC Sequences ABG61551-ABG61590 represent high growth methanotrophic  
CC bacterial strain proteins of the invention.

XX Sequence 285 AA;

Query Match 100.0%; Score 1516; DB 23; Length 285;

Best Local Similarity 100.0%; Pred. No. 1.3e-155;

Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDYAAAGGGERPAPAKLMLRLITGRPPGYHLQTVFQMLDLCMTLTFHPVVDGRVTLR 60  
DB 1 MDYAAAGGGERPAPAKLMLRLITGRPPGYHLQTVFQMLDLCMTLTFHPVVDGRVTLR 60  
QY 61 NPISGVPEQDDLTVRANLILKSHTCVGRVCIDIEKNLPMGGGLGSGSSDAATTLVLNLR 120  
DB 61 NPISGVPEQDDLTVRANLILKSHTCVGRVCIDIEKNLPMGGGLGSGSSDAATTLVLNLR 120  
QY 121 LMGLGSKRELMDLGLRLGADVVFVFGCSAMGEGVEDLQATLTPGQWFTIIRPDCHVN 180  
DB 121 LMGLGSKRELMDLGLRLGADVVFVFGCSAMGEGVEDLQATLTPGQWFTIIRPDCHVN 180  
QY 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVDAIDALCYAERLTGT 240  
DB 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVDAIDALCYAERLTGT 240  
QY 241 GACVFAQFCNKEDAESALEGLKDRWLVLFLAKGLNQSALYKYLEOG 285  
DB 241 GACVFAQFCNKEDAESALEGLKDRWLVLFLAKGLNQSALYKYLEOG 285

RESULT 2

AAE22304

ID AAE22304 standard; Protein; 285 AA.

XX AAE22304;

XX 25-JUL-2002 (first entry)

XX Methylomonas 16a sp. 4-diphosphocytidyl-2-C-methylerythritol kinase.

XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;

XX anti-oxidant; steroid; flavour; fragrance; electro-optic application;

XX aquaculture; enzyme; 4-diphosphocytidyl-2-C-methylerythritol kinase;

XX IspB.

XX Methylomonas 16a sp.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US27420.

XX 01-SEP-2000; 2000US-229858P.

XX 01-SEP-2000; 2000US-229807P.

XX (DUPLO ) DU PONT DE NEMOURS & CO E. I.

XX Brostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odum JM, Plicataggio SK, Rouviere PE;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35502.

XX Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by

XX using microorganisms having a nucleic acid molecule encoding enzymes in

XX the carotenoid biosynthetic pathway and which metabolize single carbon

XX substrates

XX Claim 44; Page 116-117; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds.  
CC The method comprises a transformed metabolising host cell, comprising  
CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule  
CC encoding an enzyme in the carotenoid biosynthetic pathway, under the  
CC control of regulatory sequences; and contacting the host cell with carbon  
CC substrate to produce a carotenoid compound. The method is useful for  
CC producing carotenoid compounds such as anthraxanthin and astaxanthin, by  
CC using microorganisms having a nucleic acid molecule encoding enzymes in  
CC the carotenoid biosynthetic pathway and which metabolize single carbon  
CC substrates. The carotenoids have potent anti-oxidant properties useful in  
CC diet, and aquaculture elements. The carotenoids are also useful as  
CC intermediates in the synthesis of steroids flavours and fragrances and  
CC compounds for potential electro-optic applications. The present sequence  
CC is Methylomonas 16a sp. 4-diphosphocytidyl-2-C-methylerythritol kinase  
CC (IspB) enzyme used in the invention.

XX Sequence 285 AA;

Query Match 100.0%; Score 1516; DB 23; Length 285;

Best Local Similarity 100.0%; Pred. No. 1.3e-155;

Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDYAAAGGGERPAPAKLMLRLITGRPPGYHLQTVFQMLDLCMTLTFHPVVDGRVTLR 60  
DB 1 MDYAAAGGGERPAPAKLMLRLITGRPPGYHLQTVFQMLDLCMTLTFHPVVDGRVTLR 60  
QY 61 NPISGVPEQDDLTVRANLILKSHTCVGRVCIDIEKNLPMGGGLGSGSSDAATTLVLNLR 120  
DB 61 NPISGVPEQDDLTVRANLILKSHTCVGRVCIDIEKNLPMGGGLGSGSSDAATTLVLNLR 120  
QY 121 LMGLGSKRELMDLGLRLGADVVFVFGCSAMGEGVEDLQATLTPGQWFTIIRPDCHVN 180  
DB 121 LMGLGSKRELMDLGLRLGADVVFVFGCSAMGEGVEDLQATLTPGQWFTIIRPDCHVN 180  
QY 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVDAIDALCYAERLTGT 240  
DB 181 TGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVVCCKLYRPVDAIDALCYAERLTGT 240  
QY 241 GACVFAQFCNKEDAESALEGLKDRWLVLFLAKGLNQSALYKYLEOG 285  
DB 241 GACVFAQFCNKEDAESALEGLKDRWLVLFLAKGLNQSALYKYLEOG 285

RESULT 3

AAU80328

ID AAU80328 standard; Protein; 285 AA.

XX AAU80328;



AA11374  
ID AAB1374 standard; protein; 283 AA.

AC AAB1374;

XX 22-FEB-2001 (first entry)

XX B. coli YCHB protein.

DE YJEB; KOTB; YOGF; YHBC; YGBB; YCHB; antibacterial;  
treatment; infection.

XX Escherichia coli.

OS DE19916176-A1.

XX 12-OCT-2000.

XX 10-APR-1999; 99DE-1016176.

XX 10-APR-1999; 99DE-1016176.

XX (FARB) BAYER AG.

XX Broetz H, Ehler K, Freidberg C, Spallmann F, Wieland B;

XX Labischinski H;

XX WPI; 2000-639611/62.

XX N-PSDB; AAC66044.

XX Essential genes from bacteria, useful in screening for antimicrobial  
agents, and related proteins, transformants and antisense sequences -

XX Disclosure; Page 23-24; 28pp; German.

CC This invention describes novel *Escherichia coli* genes (i) encoding  
proteins (iii) designated YOGF, YHBC, YGSI, YGBB, YCHB, YJEB and  
KOTB, and genes (ia) that encode orthologous gene products (iia) in  
other microorganisms and which have antibacterial activity. Recombinant  
microorganisms in which expression of (i) or (ia) can be regulated are  
used to identify compounds that bind to the gene products, particularly  
in affinity selection assays. (ii) and (iia) are used to identify, or  
prepare, antibodies and other proteins that bind to the gene products.  
Substances that bind to (ii) or (iia) are potentially useful as  
antibacterials for treating a wide range of infections in humans and  
animals. Sequences antisense to (i) and (ia) can also be used as  
antibacterials. The specified genes are widely distributed in bacteria  
but have no close homologs in eukaryotic cells.

XX Sequence 283 AA;

Query Match 40.6%; Score 616; DB 21; Length 283;  
Best Local Similarity 47.9%; Pred. No. 5.3e-58;

Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;

10 RMPAPAKLNMRLITGRBPDVGHILQTFQMLDLCDFTHFPVDDGRVTLNFIISGVPEQ 69

4 QMPSPAKLNLFVITGRBPDVGHILQTFQMLDLCDFTHFPVDDGRVTLNFIISGVPEQ 63

70 DDLTVRAANL-----KSHTCVRCVCIIDENKLMFGGGLGGSSDAATLVNLR 121

64 DNLIVRAARLMTKTAADSGRLPTG--SGANISIDKRLPMGGGLGGSSNAATLVNLR 121

122 MGIGLSKREIMDLGRLGADVVPVFGCSAMGSEVSEDLQATILPEQFVILIKPDCHVNT 181

122 WQCGLSMDLAEWGLTGLGADVVPVFGCHAFAEGVEILITPDPEKXVLVAHAGVS IPT 181

182 GEIFSSENTLNSAVVTMSDFLAGNRRDCEVVCCKLYRPVKADIDALLCYAERLTGTG 241

182 PVIRKDPFLPNTPKRSIETLLKCFSPNDCEVIARKRREVDVATLSMLLEVAERLTGTG 241

242 ACVPAQPCNKEDASALEGLKDRNL-VFLAGLNQSAALYKTL 282

DB 242 ACVPAQPCNKEDASALEGLKDRNL-WINGFVAKGANLSPLHRAM 282

RESULT 6

ID AAB68286 standard; protein; 283 AA.

XX AAB68286;

XX 09-JUL-2001 (first entry)

XX Amino acid sequence of an isopentenyl monophosphate kinase.

XX Isopentenyl monophosphate kinase; IPK; isoprenoid biosynthesis; pigment;

XX vitamin; essential oil; pathogen resistance; antibiotic; herbicide;

XX antimetabolite.

XX Escherichia coli.

XX MO200132907-A1.

XX 10-MAY-2001.

XX 02-NOV-2000; 2000MO-US30289.

XX 04-NOV-1999; 99US-0434774.

XX (UNIV) UNIV WASHINGTON STATE RES FOUND.

XX Croteau RB, Lange BM,

XX WPI; 2001-308747/32.

XX N-PSDB; AAF85079.

XX New nucleic acid encoding isopentenyl monophosphate kinase, useful e.g.  
for making transgenic plants with increased synthesis of isoprenoids,

XX e.g. essential oils -

XX Example 1; Page 48-49; 62pp; English.

CC The present sequence represents an isopentenyl monophosphate kinase (IPK)  
polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The  
IPK polynucleotide is useful recombinant production of IPK, as a source  
of probes, primers and antisense sequences and for increasing/reducing  
expression levels of IPK in cells, particularly of essential oil  
plants, so as to increase flow through the isoprenoid biosynthesis  
pathway, resulting in increased production of e.g. pigments, vitamins  
and essential oils, also to increase resistance to pests and pathogens  
CC and to improve plant strength. Mutant forms of the IPK polynucleotide  
can be used to express forms of IPK that are resistant to IPK-targeted  
herbicides, and recombinant IPK can be used to screen for antibiotics,  
herbicides and antimetabolite agents directed against IPK.

XX Sequence 283 AA;

Query Match 40.6%; Score 616; DB 22; Length 283;  
Best Local Similarity 47.9%; Pred. No. 5.3e-58;

Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;

10 RMPAPAKLNMRLITGRBPDVGHILQTFQMLDLCDFTHFPVDDGRVTLNFIISGVPEQ 69

4 QMPSPAKLNLFVITGRBPDVGHILQTFQMLDLCDFTHFPVDDGRVTLNFIISGVPEQ 63

70 DDLTVRAANL-----KSHTCVRCVCIIDENKLMFGGGLGGSSDAATLVNLR 121

64 DNLIVRAARLMTKTAADSGRLPTG--SGANISIDKRLPMGGGLGGSSNAATLVNLR 121

122 MGIGLSKREIMDLGRLGADVVPVFGCSAMGSEVSEDLQATILPEQFVILIKPDCHVNT 181

122 WQCGLSMDLAEWGLTGLGADVVPVFGCHAFAEGVEILITPDPEKXVLVAHAGVS IPT 181

182 GEIFSSENTLNSAVVTMSDFLAGNRRDCEVVCCKLYRPVKADIDALLCYAERLTGTG 241



DB 182 PVIFDPELPRNTPKRSIETLLKCEFSNDCEVIARFREYDVAVLMLLEYAPSLNLTGTG 241

QY 242 ACVFAQFCNKEDASALBGLDRMC-VPLARGLNQSAIYKEL 282

DB 242 ACVFAEPDTESEARQVLEQAPR-WLNGFVAKGANSIPLHRAK 282

RESULT 7

ID ABP25586 standard; Protein; 283 AA.

XX ABP25586;

XX ABP25586;

DT 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 348.

DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

XX group A streptococcus; Streptococcus pyogenes; antibacterial;

XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

XX MO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

PI Telford J, Masigiani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tetteilin H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN66217.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein.

PS Claim 1; Page 3188; 452536p; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX Sequence 283 AA;

Query Match 21.9%; Score 332; DB 23; Length 283;

Best Local Similarity 32.2%; Pred. No. 3, 4e-27;

Matches 85; Conservative 58; Mismatches 107; Indels 14; Gaps 5;

QY 13 AFAKINLMIRITGRBDDGHLQLQTVFQMLDLCMLTFHPVDGVTLLNPISGVPEODDL 72

DB 7 AFAKINLMIRITGRBDDGHLQLQTVFQMLDLCMLTFHPVDGVTLLNPISGVPEODDL 66

QY 73 TV-PAANLLKSHGCVAGVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSREL 131

DB 67 DVEFKAADIINKQYGINNGVHRIEKSIPVCAIGGGSTDAATIRALNRLMNLQMDYDEM 126

QY 132 MDGLREAGADVPRVPCCSAMGCVSDDQAT-TLPEQWVITIKPDCHVTGTEF--SA 187

DB 127 VAIGFKISDVPICLGCCSLVJKEGIVKPLPTLRPCWLVLPDGGISTKSTFRIDIC 186

QY 188 ENLIR-----NSAVVMSDPLAGDNEMDCSEYVCKLYREKDAIDALCYA--EARLT 238

DB 187 KSISRVDIDLKSAIISDYQLVWKSMGNSLEDITTKRNVISITIKRMLNSGADVALMT 246

QY 239 GTGACVPAQFCNKEDASALBGLK 262

DB 247 GSGPTVFSMCTEKADRVNSMK 270

RESULT 8

ID ABBS0111 standard; Protein; 293 AA.

XX ABBS0111;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #2815.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

XX vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX MO200177335-A2.

XX 13-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

XX Buchrieser C, Frangeul L, Couve E, Ruhnock C, Feihl H, Dehoux P;

PI Dussanget O, Cheouani F, Nedjari H, Glaser P, Kunst F, Cossart P;

PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Doanann E, Hain T, Berche P, Chardit A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Rose M, Voss H;

XX WPI; 2002-010914/01.

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

PT and prevention of Listeria and related bacterial infections, and

PT related polypeptides.

PS Claim 6; SEQ ID NO 2816; 192pp; French.

XX The present invention relates to the genome sequence of Listeria

CC monocytogenes BDD-e (see ABA03041). The genome sequence and fragments of

CC it are useful for selecting probes and primers for detecting genes in L.

CC polymorphisms and other genomes. The present sequence is a protein

CC encoded by the genome sequence of the present invention. Proteins

CC expressed from the genome sequence are useful for raising specific

CC antibodies, identification of L. monocytogenes and related organisms, and

CC for biosynthesis and biodegradation, especially biosynthesis of vitamin

CC B12. The genome sequence and proteins encoded by it are also useful for

CC selecting compounds that regulate gene expression and cell replication



XX The present invention relates to novel *Rhodococcus erythropolis* AM2  
 CC strain isoprenoid biosynthetic enzymes and polynucleotides encoding  
 CC such proteins. Sequences of the invention are useful for regulating  
 CC isoprenoid biosynthesis in an organism. They are useful for producing  
 CC isoprenoid compounds and gene products having enhanced or altered  
 CC activity. The present sequence is *Rhodococcus erythropolis* y9bp  
 CC (isPB) protein.

SQ Sequence 311 AA;

Query Match 18.3%; Score 278; DB 24; Length 311;  
 Best Local Similarity 44.1%; Pred. No. 2, 8e-21;

Matches 67; Conservative 22; Mismatches 57; Indels 6; Gaps 3;

QY 13 APAKIMLMRLTGRREDGYHLQTVFQMLDLCWLTFFHPDDGRVTLR---NPISGVP-E 68  
 DB 13 ASKSNVHLAVGDLRDDGYHLLTVFQALSLADTVAVPAD--TLTVRIIGDDAAAVPTD 70

QY 69 QDDLTFRANLILKSHRGVCRCVCIDIEKNLPMGGGLGGSSDAATTLVYLNRLMGLGLSK 128  
 DB 71 RNLVWRAEMMLAEEGVAPNVEIYIEKGI PVAGGAGGSSADAAALVALNSLMGLDFSR 130

QY 129 RELMDLGLRLGADVPVFFVCGSAMGEVSEDL 160  
 DB 131 PDLDAFRAALGSDVPSLHGTLALGTGRGEQL 162

RESULT 11

AA68287  
 ID AAB68287 standard; Protein; 401 AA.

AC AAB68287;

DT 09-JUL-2001 (first entry)

XX Amino acid sequence of an isopentenyl monophosphatase kinase.

XX Isopentenyl monophosphatase kinase; IPK; isoprenoid biosynthesis; pigment;  
 KM vitamin; essential oil; pathogen resistance; antibiotic; herbicide;  
 KM antimutagenic.

XX Lycopersicon esculentum.

PN M0200132907-A1.

PD 10-MAY-2001.

PF 02-NOV-2000; 2000WO-US30289.

PR 04-NOV-1999; 99US-0434774.

PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.

PI Croteau RB, Lange BM;

DR WPI; 2001-308747/32.

DR N-PSDB; AAF85082.

XX New nucleic acid encoding isopentenyl monophosphatase kinase, useful e.g.  
 PT for making transgenic plants with increased synthesis of isoprenoids,  
 PT e.g. essential oils

PS Example 8; Page 52-53; 62pp; English.

XX The present sequence represents an isopentenyl monophosphatase kinase (IPK)  
 CC polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The  
 CC IPK polynucleotide is useful recombinant production of IPK, as a source  
 CC of probes, primers and antisense sequences and for increasing/reducing  
 CC expression levels of IPK in cells, particularly of essential oil  
 CC plants, so as to increase flow through the isoprenoid biosynthesis  
 CC pathway, resulting in increased production of e.g. pigments, vitamins  
 CC and essential oils, also to increase resistance to pests and pathogens

CC and to improve plant strength. Mutant forms of the IPK polynucleotide  
 CC can be used to express forms of IPK that are resistant to IPK-targeted  
 CC herbicides, and recombinant IPK can be used to screen for antibiotics,  
 CC herbicides and antimutagenic agents directed against IPK.

SQ Sequence 401 AA;

Query Match 18.3%; Score 277; DB 22; Length 401;  
 Best Local Similarity 34.0%; Pred. No. 5, 1e-21;

Matches 71; Conservative 43; Mismatches 89; Indels 6; Gaps 4;

QY 13 APAKIMLMRLTGRREDGYHLQTVFQMLDLCWLTFFHPDDGRVTLRNPISGVP-E 70  
 DB 93 SPCKINFLRTSRKRDGYHLLSLFVHISLGDKIKSLSPSKSRISTVAVGPIDER 152

QY 71 DLTFRANLILKSHRGVCRCVCIDIEKNLPMGGGLGGSSDAATTLVYLNRLMGLGLSKRE 130  
 DB 153 NLTKALMLRKTKTGTRVFIHLDKVPFGAGLGGSSNAATTLMAANQSGCAVATEKE 212

QY 131 LMDLGLRLGADVPVFFVCGSAMGEVSEDLQATLTPBQW---PVYIKPDCHVNTGELFSA 187  
 DB 213 LQWSEIGSDIPFPFSGHAAVCTGRGEVVDIPSPIDIPVTLIKPOACSTAEVYKR 272

QY 188 ENLFRNSAVVTMSDFLAGDNRCSEVVC 216  
 DB 273 FQDLSSKVDPLS-LLEKISTGIGSDVC 300

RESULT 12

AA68288  
 ID AAB68288 standard; Protein; 383 AA.

AC AAB68288;

DT 09-JUL-2001 (first entry)

XX Amino acid sequence of an isopentenyl monophosphatase kinase.

XX Isopentenyl monophosphatase kinase; IPK; isoprenoid biosynthesis; pigment;  
 KM vitamin; essential oil; pathogen resistance; antibiotic; herbicide;  
 KM antimutagenic.

XX Arabidopsis thaliana.

PN M0200132907-A1.

PD 10-MAY-2001.

PF 02-NOV-2000; 2000WO-US30289.

PR 04-NOV-1999; 99US-0434774.

PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.

PI Croteau RB, Lange BM;

DR WPI; 2001-308747/32.

DR N-PSDB; AAF85083.

XX New nucleic acid encoding isopentenyl monophosphatase kinase, useful e.g.  
 PT for making transgenic plants with increased synthesis of isoprenoids,  
 PT e.g. essential oils

PS Example 8; Page 56-57; 62pp; English.

XX The present sequence represents an isopentenyl monophosphatase kinase (IPK)  
 CC polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The  
 CC IPK polynucleotide is useful recombinant production of IPK, as a source  
 CC of probes, primers and antisense sequences and for increasing/reducing  
 CC expression levels of IPK in cells, particularly of essential oil  
 CC plants, so as to increase flow through the isoprenoid biosynthesis  
 CC pathway, resulting in increased production of e.g. pigments, vitamins  
 CC and essential oils, also to increase resistance to pests and pathogens

CC and to improve plant strength. Mutant forms of the IPK polynucleotide  
CC can be used to express forms of IPK that are resistant to IPK-targeted  
CC herbicides, and recombinant IPK can be used to screen for antibiotics,  
CC herbicides and antimetabolic agents directed against IPK.

XX Sequence 383 AA;

Query Match 18.1%; Score 275; DB 22; Length 383;  
Best Local Similarity 33.5%; Pred. No. 7,9e-21;  
Matches 70; Conservative 42; Mismatches 91; Indels 6; Gaps 4;

QY 13 AAKAKMLKRTGRPDGTHLLQTVFQMLDLCMLTF--PVDDGRVTL-RNPISGVPEQ 70  
DB 77 SCKKINVFRTIGKEDGDFHDLASLFHVISLGDITKPSISPKSKDRISTVQGVVQGR 136  
QY 71 DLTVRAANLLKSHTCGVRGVCIDIEKNLPMGGGLOGSSDAATTVLVNRLMGLSKRE 130  
DB 137 NLIIRKLNLYRKKTGSNRFWHLDKVPTGAGLOGSSNATLMAANLNGGLVTENE 196  
QY 131 LMDLGLRLGADVPVFGCSAMGEVSEDLQATLP--EQWVTLIKPDCHVNGEIFSA 187  
DB 197 LQDWSELSISDIPFPFSHGAAYCTGRGEIVQDLPPFPDLPMVLKPREACSTAVYKR 256  
QY 188 ENLTRSAVAVTMSDFLAGDNRDCEVVC 216  
DB 257 LRLDQTSNINPLT-LLENVTSNGVSGSTC 284

# RESULT 13

AAB68291  
ID AAB68291 standard; Protein; 315 AA.

AC AAB68291;  
DT 09-JUL-2001 (first entry)

DE Amino acid sequence of an isopentenyl monophosphatase kinase.

XX Isopentenyl monophosphatase kinase; IPK; isoprenoid biosynthesis; pigment;  
KM vitamin; essential oil; pathogen resistance; antibiotic; herbicide;  
KW antimetabolic.

OS Synchocystis sp.

PN MO200132907-A1.

PD 10-MAY-2001.

PF 02-NOV-2000; 2000MO-US30289.

PR 04-NOV-1999; 99US-0434774.

XX (UNITW ) UNITV WASHINGTON STATE RES FOUND.

PI Croteau RB, Lange EM;

DR WPI; 2001-308747/32.

XX N-PSDB; AAF85084.

PT New nucleic acid encoding isopentenyl monophosphatase kinase, useful e.g.  
PT for making transgenic plants with increased synthesis of isoprenoids,  
PT e.g. essential oils

PS Example 8; Page 60-61; 62pp; English.

CC The present sequence represents an isopentenyl monophosphatase kinase (IPK)  
CC polypeptide. The enzyme is an isoprenoid biosynthesis stimulator. The  
CC IPK polynucleotide is useful recombinant production of IPK, as a source  
CC of probes, primers and antisense sequences and for increasing/reducing  
CC expression levels of IPK in cells, particularly of essential oil  
CC plants, so as to increase flow through the isoprenoid biosynthesis  
CC pathway, resulting in increased production of e.g. pigments, vitamins  
CC and essential oils, also to increase resistance to pests and pathogens

CC and to improve plant strength. Mutant forms of the IPK polynucleotide  
CC can be used to express forms of IPK that are resistant to IPK-targeted  
CC herbicides, and recombinant IPK can be used to screen for antibiotics,  
CC herbicides and antimetabolic agents directed against IPK.

XX Sequence 315 AA;

Query Match 17.8%; Score 270.5; DB 22; Length 315;  
Best Local Similarity 30.3%; Pred. No. 1.8e-20;  
Matches 91; Conservative 49; Mismatches 109; Indels 51; Gaps 12;

QY 13 AAKAKMLKRTGRPDGTHLLQTVFQMLDLCMLTF--HPVDDGRVTL-RNPISGVPEQ 69  
DB 8 AFAKINFLBETIGRPPGFHFLVWVLOSIALGDKITVRANGTDIRLSCDPSPLAN--DA 65  
QY 70 DDLTVRAANLLKSHTCGVRGVCIDIEKNLPMGGGLOGSSDAATTVLVNRLMGL 124  
DB 66 TNLVRAANLQMINNPPQAHN-YGVVITLTKHLPMAAGLAGSADAAVAVYGLDLMNL 124  
QY 125 GLSKRELMDGLRLGADVPVFGCSAMGEVSEDLQATLP--WVTLIK-PDCHVN 180  
DB 125 GLTRPELEQLAAQLGSDIPFCIGGTAIATGRGILD--FLPDGNCVAVLAKRSIEVS 182  
QY 181 T-----GEIFAEINLTRSAVVTM-----SDPLAGDNRDCEVVC 216  
DB 183 TPMAVQTRQKFGKXVYNDQSQARRKTHAGPLGIGIHRNPGQIASHIHNDLEKVL 242  
QY 217 KLYRPVDAIDALCYA--EARLTGCAVFAQPCNREDALEAGLDR-----WE 266  
DB 243 PAHQVPAQLRQVLQSAGSLGTPWMSGSPSVFTLREQAEMQVLAIAKEXLNDPVDVWL 302

# RESULT 14

AAB70486  
ID AAB70486 standard; Protein; 316 AA.

AC AAB70486;

DT 04-MAY-2001 (first entry)

DE A. thaliana ychb protein without leader sequence.

XX Isoprenoid biosynthesis; 2C-methyl-D-erythritol-4-phosphate; herbicide;  
KM genetic marker; plant; isoprenoid biosynthesis inhibitor.

OS Arabidopsis thaliana.

PN MO200111055-A1.

PD 15-FEB-2001.

PF 03-AUG-2000; 2000MO-EP07548.

PR 04-AUG-1999; 99DB-1036663.

PR 21-SEP-1999; 99DB-1045174.

PR 21-SEP-1999; 99DB-1045175.

PR 11-OCT-1999; 99DB-1046887.

PR 05-NOV-1999; 99DB-1053309.

PR 28-APR-2000; 2000DB-1020996.

XX (BACH/) BACHER A.

PA (ZENK/) ZENK M.

PI Bacher A, Zenk M, Eisenreich W, Fellermeier M, Fischer M, Hecht S,

PI Herz S, Kis K, Luettgen H, Rohdich F, Sagner S, Schuhr CA,  
PI Kungsintaweekul J;  
DR WPI; 2001-202774/20.  
XX N-PSDB; AAF63675.

PT Enzymes which operate in the alternative isoprenoid pathway downstream  
PT from 2C-methyl-D-erythritol-4-phosphate, useful for screening a  
PT chemical library for inhibitors of the biosynthesis of isoprenoids -

XX Example 12a; Annex Ea; 194pp; English.

XX The present invention describes functional enzymes (I) which operate in  
 CC the alternative isoprenoid pathway downstream from  
 CC 2C-methyl-D-erythritol-4-phosphate. (I) can be used as isoprenoid  
 CC biosynthesis inhibitor. The enzymes are useful for screening a chemical  
 CC library for inhibitors of the biosynthesis of isoprenoids.  
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol,  
 CC 2C-methyl-D-erythritol-2,4-cyclopyrophosphate or  
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate, or their salts  
 CC are useful for screening for inhibitors of the biosynthesis of  
 CC isoprenoids. Herbicide resistant enzymes can be used as genetic markers  
 CC in any cell that is normally sensitive to the inhibitory effects of  
 CC herbicide formation. Methods from the present invention can be used to  
 CC produce herbicide resistant enzyme variants which can be incorporated  
 CC into plants to confer selective herbicide resistant on the plants. The  
 CC methods are also useful for controlling weed by cultivating crops  
 CC containing herbicide-resistant genes in the presence of weed-controlling  
 CC effective amounts of herbicides. The inhibitors are used for inhibiting  
 CC the biosynthesis of isoprenoids in plants, bacteria or protozoa. The  
 CC present sequence represents the Arabidopsis thaliana ychb protein without  
 CC leader sequence, which is given in an example from the present invention.

XX Sequence 316 AA;

Query Match 17.7%; Score 268; DB 22; Length 316;  
 Best Local Similarity 33.0%; Pred. No. 3.5e-20;  
 Matches 69; Conservative 42; Mismatches 92; Indels 6; Gaps 4;

QY 13 ABAKNTMLRTGRPPDGYHLLQTVFQMLDCMWTFR-EVDDGRVTLKPNISGVP-EQD 70  
 DB 10 SPCKINVFRLITGRKEDGFHDLALFHVLSIGTIKFSLSPSKSDRLSTWQGVVDGR 69  
 QY 71 DLTVAANILKSHNGCVRGVCIDIEKNLPMGGGAGGSSDATTLVNRLMGGLSKRE 130  
 DB 70 NLTITLNLVKKTKGSRNRFPHILDKVPTGAGLGGSSNATLMAANLNGGLVENE 129  
 QY 131 LMDLGLRLGADVVPVFGCSAMGEGVSEDLQATLP--EQWFIILKPDCHVNTGEISFA 187  
 DB 130 LQWSSSEIGSDIPFPFSGAAYCTGREGIVQDLPPPPFLDPMWLKPRBASTAEVYKR 189  
 QY 188 ENLIRNSAVVTMSDFLAGDNRDCEVVC 216  
 DB 190 ERLDQTSNINPLT-LIKVNTSGVQSIC 217

Result 15

AAE70487  
 ID AAB70487 standard; Protein; 383 AA.

AC AAB70487;

DT 04-MAY-2001 (first entry)

DE Arabidopsis thaliana ychb protein.

XX Isoprenoid biosynthesis; 2C-methyl-D-erythritol-4-phosphate; herbicide;  
 KM genetic marker; plant; isoprenoid biosynthesis inhibitor.

XX Arabidopsis thaliana.

OS WO200111055-A1.

PN 15-FEB-2001.

PF 03-AUG-2000; 2000WO-EP07548.

PR 04-AUG-1999; 99DE-1036663.

PR 21-SEP-1999; 99DE-1045174.

PR 11-OCT-1999; 99DE-1045175.

PR 05-NOV-1999; 99DE-1053309.

PR 28-APR-2000; 2000DE-1020996.

PA (BACH/) BACHER A.  
 PA (ZENK/) ZENK M.

PI Bacher A, Zenk M, Eisenreich W, Fellermeier M, Fischer M, Hecht S;  
 PI Herz S, Kis K, Luettgen H, Rohdich F, Sagner S, Schuhr CA;

PI Wugsintaweekul J;

DR WPI; 2001-202774/20.

DR N-PSDB; AAF63680.

PT Enzymes which operate in the alternative isoprenoid pathway downstream  
 PT from 2C-methyl-D-erythritol-4-phosphate, useful for screening a  
 PT chemical library for inhibitors of the biosynthesis of isoprenoids -

PS Example 12b; Annex Eb; 194pp; English.

XX The present invention describes functional enzymes (I) which operate in  
 CC the alternative isoprenoid pathway downstream from  
 CC 2C-methyl-D-erythritol-4-phosphate. (I) can be used as isoprenoid  
 CC biosynthesis inhibitor. The enzymes are useful for screening a chemical  
 CC library for inhibitors of the biosynthesis of isoprenoids.  
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol,  
 CC 2C-methyl-D-erythritol-2,4-cyclopyrophosphate or  
 CC 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate, or their salts  
 CC are useful for screening for inhibitors of the biosynthesis of  
 CC isoprenoids. Herbicide resistant enzymes can be used as genetic markers  
 CC in any cell that is normally sensitive to the inhibitory effects of  
 CC herbicide formation. Methods from the present invention can be used to  
 CC produce herbicide resistant enzyme variants which can be incorporated  
 CC into plants to confer selective herbicide resistant on the plants. The  
 CC methods are also useful for controlling weed by cultivating crops  
 CC containing herbicide-resistant genes in the presence of weed-controlling  
 CC effective amounts of herbicides. The inhibitors are used for inhibiting  
 CC the biosynthesis of isoprenoids in plants, bacteria or protozoa. The  
 CC present sequence represents the Arabidopsis thaliana ychb protein.  
 CC which is given in an example from the present invention.

XX Sequence 383 AA;

Query Match 17.7%; Score 268; DB 22; Length 383;  
 Best Local Similarity 33.0%; Pred. No. 4.6e-20;  
 Matches 69; Conservative 42; Mismatches 92; Indels 6; Gaps 4;

QY 13 ABAKNTMLRTGRPPDGYHLLQTVFQMLDCMWTFR-EVDDGRVTLKPNISGVP-EQD 70

DB 77 SPCKINVFRLITGRKEDGFHDLALFHVLSIGTIKFSLSPSKSDRLSTWQGVVDGR 136

QY 71 DLTVAANILKSHNGCVRGVCIDIEKNLPMGGGAGGSSDATTLVNRLMGGLSKRE 130

DB 137 NLTITLNLVKKTKGSRNRFPHILDKVPTGAGLGGSSNATLMAANLNGGLVENE 196

QY 131 LMDLGLRLGADVVPVFGCSAMGEGVSEDLQATLP--EQWFIILKPDCHVNTGEISFA 187

DB 197 LQWSSSEIGSDIPFPFSGAAYCTGREGIVQDLPPPPFLDPMWLKPRBASTAEVYKR 256

QY 188 ENLIRNSAVVTMSDFLAGDNRDCEVVC 216

DB 257 ERLDQTSNINPLT-LIKVNTSGVQSIC 284

Search completed: January 29, 2004, 15:49:42  
 Job time : 28.0514 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:45:34 : Search time 8.94668 Seconds  
(without alignments)  
1346.130 Million cell updates/sec

Title: US-09-941-947A-12  
Perfect score: 1516  
Sequence: 1 MDYAAGKGRWPAAPALNLM.....LVFLAKGLNQSALYKKLEGG 285

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5  
Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgm2\_6/prodata/1/iaa/5A\_COMB.pep.\*  
2: /cgm2\_6/prodata/1/iaa/5B\_COMB.pep.\*  
3: /cgm2\_6/prodata/1/iaa/6A\_COMB.pep.\*  
4: /cgm2\_6/prodata/1/iaa/6B\_COMB.pep.\*  
5: /cgm2\_6/prodata/1/iaa/6C\_COMB.pep.\*  
6: /cgm2\_6/prodata/1/iaa/backfillseq1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID                     | Description        |
|------------|-------|-------------|--------|------------------------|--------------------|
| 1          | 668   | 44.1        | 294    | 4 US-09-252-991A-30806 | Sequence 30806, A  |
| 2          | 616   | 40.6        | 283    | 3 US-09-434-774-6      | Sequence 6, Appli  |
| 3          | 589.5 | 38.9        | 283    | 4 US-09-328-352-4670   | Sequence 4670, Ap  |
| 4          | 302   | 19.9        | 289    | 4 US-09-134-001C-3348  | Sequence 3348, Ap  |
| 5          | 277   | 18.3        | 401    | 2 US-08-596-111B-2     | Sequence 2, Appli  |
| 6          | 277   | 18.3        | 401    | 3 US-09-434-774-10     | Sequence 10, Appl  |
| 7          | 275   | 18.1        | 383    | 3 US-09-434-774-12     | Sequence 12, Appl  |
| 8          | 270.5 | 17.8        | 315    | 3 US-09-434-774-16     | Sequence 16, Appl  |
| 9          | 263   | 17.3        | 405    | 3 US-09-434-774-2      | Sequence 2, Appli  |
| 10         | 178   | 11.7        | 141    | 4 US-09-198-452A-1236  | Sequence 1236, Ap  |
| 11         | 160   | 10.6        | 99     | 4 US-09-634-238-305    | Sequence 305, App  |
| 12         | 91.5  | 6.0         | 783    | 1 US-08-443-521-2      | Sequence 2, Appli  |
| 13         | 91.5  | 6.0         | 783    | 3 US-09-012-871-2      | Sequence 4141, Ap  |
| 14         | 89    | 5.9         | 315    | 4 US-09-134-001C-4141  | Sequence 4141, Ap  |
| 15         | 86    | 5.7         | 312    | 4 US-09-252-991A-24594 | Sequence 24594, A  |
| 16         | 84.5  | 5.6         | 243    | 4 US-08-588-207A-338   | Sequence 338, App  |
| 17         | 84    | 5.5         | 793    | 4 US-09-328-352-4283   | Sequence 4283, Ap  |
| 18         | 81.5  | 5.4         | 3052   | 2 US-08-557-122A-26    | Sequence 26, Appl  |
| 19         | 81.5  | 5.4         | 3052   | 4 US-09-262-666-26     | Sequence 26, Appl  |
| 20         | 78.5  | 5.3         | 1118   | 4 US-09-585-173B-36    | Sequence 36, Appl  |
| 21         | 78    | 5.1         | 324    | 4 US-09-634-238-307    | Sequence 307, Appl |
| 22         | 78    | 5.1         | 324    | 3 US-08-874-347-20     | Sequence 20, Appl  |
| 23         | 77    | 5.1         | 518    | 4 US-09-093-522-20     | Sequence 20, Appl  |
| 24         | 77    | 5.1         | 518    | 4 US-09-043-123-4      | Sequence 4, Appli  |
| 25         | 77    | 5.1         | 518    | 4 US-08-973-398-2      | Sequence 2, Appli  |
| 26         | 77    | 5.1         | 518    | 5 PCT-US86-05800-7     | Sequence 7, Appli  |
| 27         | 76.5  | 5.0         | 259    | 6 5223425-2            | Patent No. 5223425 |

|    |      |     |      |                        |                    |
|----|------|-----|------|------------------------|--------------------|
| 28 | 76.5 | 5.0 | 260  | 6 5223425-10           | Patent No. 5223425 |
| 29 | 76.5 | 5.0 | 455  | 4 US-09-328-352-5626   | Sequence 5626, Ap  |
| 30 | 76.5 | 5.0 | 531  | 3 US-08-886-886-13     | Sequence 15, Appl  |
| 31 | 75.5 | 5.0 | 390  | 4 US-09-252-991A-19993 | Sequence 19993, A  |
| 32 | 75.5 | 5.0 | 472  | 3 US-08-348-518C-5     | Sequence 5, Appli  |
| 33 | 75.5 | 5.0 | 472  | 3 US-08-476-509B-5     | Sequence 5, Appli  |
| 34 | 75   | 4.9 | 392  | 1 US-08-451-777A-33    | Sequence 33, Appl  |
| 35 | 75   | 4.9 | 392  | 2 US-08-451-778A-33    | Sequence 33, Appl  |
| 36 | 75   | 4.9 | 1891 | 2 US-08-998-208-33     | Sequence 33, Appl  |
| 37 | 75   | 4.9 | 1891 | 2 US-08-804-227C-12    | Sequence 12, Appli |
| 38 | 75   | 4.9 | 1891 | 2 US-08-804-198-6      | Sequence 6, Appli  |
| 39 | 74.5 | 4.9 | 221  | 4 US-09-107-532A-5984  | Sequence 5984, Ap  |
| 40 | 74.5 | 4.9 | 397  | 4 US-09-107-532A-5568  | Sequence 5568, Ap  |
| 41 | 74.5 | 4.9 | 590  | 4 US-09-134-001C-4390  | Sequence 4390, Ap  |
| 42 | 74   | 4.9 | 462  | 4 US-09-328-352-4742   | Sequence 4742, Ap  |
| 43 | 73.5 | 4.8 | 486  | 3 US-08-348-518C-2     | Sequence 2, Appli  |
| 44 | 73.5 | 4.8 | 514  | 1 US-08-190-802A-66    | Sequence 66, Appli |
| 45 | 73.5 | 4.8 | 514  | 3 US-08-477-346-66     | Sequence 66, Appli |

ALIGNMENTS

RESULT 1  
US-09-252-991A-30806  
; Sequence 30806, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubinfeld et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30806  
; LENGTH: 294  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30806

|                       |  |   |          |            |
|-----------------------|--|---|----------|------------|
| Query Match           | 44.1%                                  | Score 668   | DB 4     | Length 294 |
| Best local Similarity | 51.8%                                  | Pred. No. 1.8e-65   |          |            |
| Matches 141           | Conservative 36                        | Mismatches 95   | Indels 0 | Gaps 0     |
| QY                    | 12                                     | PAPALNLMRLTGRRRPGYHLLQVFMLOLDCWLTFRPDDGVTLENPISGVPEQDD    | 71       |            |
| DB                    | 20                                     | PAPALNLMRLTGRRRPGYHLLQVFMLOLDCWLTFRPDDGVTLENPISGVPEQDD    | 79       |            |
| QY                    | 72                                     | LIVRAANLKSHGCVGRVCIDIEKNLPMGGGLGGSSDAAITLVNLRMLGUSKREL    | 131      |            |
| DB                    | 80                                     | LIVRAARGLQSSGSPQGVDFWMDRLPMGGIGGSSDAAITLVNLRMLGUSKREL     | 139      |            |
| QY                    | 132                                    | MDGLRLGADVVFVFGCSAMGEGVSEDLQAITLPKQFVLIKPCCHVNTGIIIFAEHLT | 191      |            |
| DB                    | 140                                    | MDGLRLGADVVFVFGCSAMGEGVSEDLQAITLPKQFVLIKPCCHVNTGIIIFAEHLT | 199      |            |
| QY                    | 192                                    | RNSAVVTMSDFIAGNRNDCSVVCKLTPYVMDAIDALICVAEALNTGACVFAQCNK   | 251      |            |
| DB                    | 200                                    | RDSFALIKRIVLEGRSRDQPVYERRRPEVRNMLILNKTVSALITGIGCVFSGFPNK  | 259      |            |
| QY                    | 252                                    | BDASALGLQDRWLVFLAKGLNQSALYKKLE                            | 283      |            |
| DB                    | 260                                    | BDASALGLQDRWLVFLAKGLNQSALYKKLE                            | 291      |            |
| RESULT 2              | US-09-434-774-6                        |   |          |            |
|                       | ; Sequence 6, Application US/09434774A |   |          |            |





STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 10-APR-1996  
CLASSIFICATION: 800  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: GB 9315751.9  
FILING DATE: 30-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: UPLA:068  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512)418-3000  
TELEFAX: (512)474-7577  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 401 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-596-111B-2

Query Match 18.3%; Score 277; DB 2; Length 401;  
Best Local Similarity 34.0%; Pred. No. 4.5e-22;  
Matches 71; Conservative 43; Mismatches 89; Indels 6; Gaps 4.

QY 13 APAKINLMRLITGRPDGVYHLLQTVFQMLDLCDWLTFH-PVDDGRVTLNRPISGVP-EOD 70  
DB 93 SPCKINVFRLITGRPDGVYHLLQTVFQMLDLCDWLTFH-PVDDGRVTLNRPISGVP-EOD 70  
QY 71 DLITRAANLLKSHGTGVRGVCIDIEKPLPMGGIGGSSDAATLVNRLMGLGSKRE 130  
DB 153 NLITKALNLYRKKTGTNDYFWIHLDKVPTGAGLGSSNNAATLTMANQSGCVATKE 212  
QY 131 LMDLGLRLGADVPVFFVFGCSAMGEGVSEDLQATLPEQW---FVLIKPDCHVNTGEIPSA 187  
DB 213 LQMSGEIGSDIPFPFSGAAYCTGREGVYQDIPSPFPDIPMLIKRQOASTAEVYKR 272  
QY 188 ENLTNSAVVTMSDFLAGDNRNDCSEVYC 216  
DB 273 FQDLSSKVDPLS-LLEKISTSGISQDVC 300

RESULT 5  
US-09-434-774-10  
Sequence 10, Application US/09434774A  
Patent No. 6235514  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B  
APPLICANT: Lange, Bernd M  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPENTENYL  
TITLE OF INVENTION: MONOPHOSPHATE KINASE, AND METHODS OF USE  
FILE REFERENCE: wsu14448  
CURRENT APPLICATION NUMBER: US/09/434,774A  
CURRENT FILING DATE: 1999-11-04  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 10  
LENGTH: 401  
TYPE: PRT  
ORGANISM: Lycopersicon esculentum  
US-09-434-774-10

Query Match 18.3%; Score 277; DB 3; Length 401;  
Best Local Similarity 34.0%; Pred. No. 4.5e-22;  
Matches 71; Conservative 43; Mismatches 89; Indels 6; Gaps 4.

QY 13 APAKINLMRLITGRPDGVYHLLQTVFQMLDLCDWLTFH-PVDDGRVTLNRPISGVP-EOD 70  
DB 93 SPCKINVFRLITGRPDGVYHLLQTVFQMLDLCDWLTFH-PVDDGRVTLNRPISGVP-EOD 70  
QY 71 DLITRAANLLKSHGTGVRGVCIDIEKPLPMGGIGGSSDAATLVNRLMGLGSKRE 130  
DB 153 NLITKALNLYRKKTGTNDYFWIHLDKVPTGAGLGSSNNAATLTMANQSGCVATKE 212  
QY 131 LMDLGLRLGADVPVFFVFGCSAMGEGVSEDLQATLPEQW---FVLIKPDCHVNTGEIPSA 187  
DB 213 LQMSGEIGSDIPFPFSGAAYCTGREGVYQDIPSPFPDIPMLIKRQOASTAEVYKR 272  
QY 188 ENLTNSAVVTMSDFLAGDNRNDCSEVYC 216  
DB 273 FQDLSSKVDPLS-LLEKISTSGISQDVC 300

RESULT 7  
US-09-434-774-12  
Sequence 12, Application US/09434774A  
Patent No. 6235514  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B  
APPLICANT: Lange, Bernd M  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPENTENYL  
TITLE OF INVENTION: MONOPHOSPHATE KINASE, AND METHODS OF USE  
FILE REFERENCE: wsu14448  
CURRENT APPLICATION NUMBER: US/09/434,774A  
CURRENT FILING DATE: 1999-11-04  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 12  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-09-434-774-12

Query Match 18.1%; Score 275; DB 3; Length 383;  
Best Local Similarity 33.5%; Pred. No. 7e-22;  
Matches 70; Conservative 42; Mismatches 91; Indels 6; Gaps 4.

QY 13 APAKINLMRLITGRPDGVYHLLQTVFQMLDLCDWLTFH-PVDDGRVTLNRPISGVP-EOD 70  
DB 77 SPCKINVFRLITGRPDGVYHLLQTVFQMLDLCDWLTFH-PVDDGRVTLNRPISGVP-EOD 70  
QY 71 DLITRAANLLKSHGTGVRGVCIDIEKPLPMGGIGGSSDAATLVNRLMGLGSKRE 130  
DB 137 NLITKALNLYRKKTGTNDYFWIHLDKVPTGAGLGSSNNAATLTMANQSGCVATKE 196  
QY 131 LMDLGLRLGADVPVFFVFGCSAMGEGVSEDLQATLPEQW---FVLIKPDCHVNTGEIPSA 187  
DB 197 LQMSGEIGSDIPFPFSGAAYCTGREGVYQDIPSPFPDIPMLIKRQOASTAEVYKR 256  
QY 188 ENLTNSAVVTMSDFLAGDNRNDCSEVYC 216  
DB 257 LRLDQTSINPLT-LLEVTNSGVQSIC 284

RESULT 8  
US-09-434-774-16  
Sequence 16, Application US/09434774A  
Patent No. 6235514  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B  
APPLICANT: Lange, Bernd M  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPENTENYL  
TITLE OF INVENTION: MONOPHOSPHATE KINASE, AND METHODS OF USE  
FILE REFERENCE: wsu14448  
CURRENT APPLICATION NUMBER: US/09/434,774A

CURRENT FILING DATE: 1999-11-04  
 NUMBER OF SEQ ID NOS: 16  
 SOFTWARE: Patent Ver. 2.0  
 SEQ ID NO 16  
 LENGTH: 315  
 TYPE: PRT  
 ORGANISM: *Synechocystis* sp.  
 US-09-434-774-16

Query Match 17.8%; Score 270.5; DB 3; Length 315;  
 Best Local Similarity 30.3%; Pred. No. 1.7e-21;  
 Matches 91; Conservative 49; Mismatches 109; Indels 51; Gaps 12;

13 APAKNTLMRTGRRPDGYHLLQTVFOMLDCMTLTF--HPVDDGVTLRNPISGVPEQ 69  
 8 APTKTLFETLADRPDGHVLMVQSLAGSKITVRANGTDDITLSCGDSPLAN--DA 65  
 70 DDLTVRAANIL-----KSHTCGVCYCIDIEKNLPMGGGLGSSSDAATTLVLRNLML 124  
 66 TNLAYRAAQULMNNFPQADHN--YGVVDITLTKHIFMAAGLAGSADAAVAVGLDULMNL 124  
 125 GUSKREIMDLGLRGLADVPPVFGCSAMGSEVSEDLQATLPEQ--WFTYIK-PDCRYN 180  
 125 GLTRPELEQLAQLGSDIPFCIGGTAITGRGELID--PLPDGNCFFWVYLAKEHSIEVS 182  
 181 T-----GEISAEMLTNSAVVTM-----SDPLAGDNNDSCSEVVC 216  
 183 TPAAYOTYOKKOKNVLNDQSRARAKTIHAGPLLOGIQRHNPQJASHINDLAKVYL 242  
 217 KLYRPVKADIALLCYA--EARITGACVFAQFCKEDAESALEGLKDR-----WL 266  
 243 PAHQPAQLRQVLOSAGLGATWMSGSGSPVFTLCREQAEAOVLAIAREKLNDDPVDFPL 302

RESULT 9  
 US-09-434-774-2  
 Sequence 2, Application US/09434774A  
 Patent No. 623514  
 GENERAL INFORMATION:  
 APPLICANT: Croteau, Rodney B  
 APPLICANT: Lange, Bernd M  
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPRENTAL  
 FILE REFERENCE: wstut14448  
 CURRENT APPLICATION NUMBER: US/09/434,774A  
 CURRENT FILING DATE: 1999-11-04  
 NUMBER OF SEQ ID NOS: 16  
 SOFTWARE: Patent Ver. 2.0  
 SEQ ID NO 2  
 LENGTH: 405  
 TYPE: PRT  
 ORGANISM: *Mentha piperita*  
 US-09-434-774-2

Query Match 17.3%; Score 263; DB 3; Length 405;  
 Best Local Similarity 30.2%; Pred. No. 1.6e-20;  
 Matches 76; Conservative 50; Mismatches 104; Indels 22; Gaps 7;

13 APAKNTLMRTGRRPDGYHLLQTVFOMLDCMTLTF--HPVDDGVTLRNPISGVPEQ 69  
 102 SPCKINVFRTGCKREDGFHDLASLEHVISLDGKIKFSLSPKFNAGFPVTN--VPGVPLDS 160  
 70 DDLTVRAANILKSHTCGVCYCIDIEKNLPMGGGLGSSSDAATTLVLRNLMLGLSKR 129  
 161 KNLITLALNLFKKKSTVDHFMHLDKVPRTAGAGGSSNATLMAANDPSSGCIATEK 220  
 130 BLMDLGLRGLADVPPVFGCSAMGSEVSEDLQATLPE--EOMFYLIKPDCEVNTGEIFS 186  
 221 DLQMSGEIGSDIPFFFSGAAYCTGRGEVEDIPPPVROLDSVIAKPYQCEACPTGEVYK 280  
 187 AENLRNSAVVTM-----SDPLAGDNNDSCSEVVCVLYRPVKADIAL--LCY 232  
 281 RLRLDQTSIDIPVLTLEKISKGGISODVCNLDLBPAPAEVVSLEK-LKQRIAAAGRSQY 339

QY 233 AEARLTGTACV 244  
 DB 340 DAVFMSGSGSTI 351

RESULT 10  
 US-09-198-452A-1236  
 Sequence 1236, Application US/09198452A  
 Patent No. 6559294  
 GENERAL INFORMATION:  
 APPLICANT: Griffais, R  
 TITLE OF INVENTION: *Chlamydia pneumoniae* genomic sequence and polypeptides, fragments  
 TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 FILE REFERENCE: 9710-003-999  
 CURRENT APPLICATION NUMBER: US/09/198,452A  
 CURRENT FILING DATE: 1998-11-24  
 NUMBER OF SEQ ID NOS: 6849  
 SEQ ID NO 1236  
 LENGTH: 141  
 TYPE: PRT  
 ORGANISM: *Chlamydia pneumoniae*  
 US-09-198-452A-1236

Query Match 11.7%; Score 178; DB 4; Length 141;  
 Best Local Similarity 32.3%; Pred. No. 8.4e-12;  
 Matches 43; Conservative 30; Mismatches 58; Indels 2; Gaps 1;

13 APAKNTLMRTGRRPDGYHLLQTVFOMLDCMTLTHPVDDGVTLRNPISGVPEQDDL 72  
 5 SPATINLEFLKTKWGRFPNFHETLTYQADIDFGDTLSK--NSMDLSLSNNNELSPSNL 62  
 73 TVRANILKSHTCGVCYCIDIEKNLPMGGGLGSSSDAATTLVLRNLMLGLSKREIM 132  
 63 TWKSLBIFRRETQHQVSWHUNKSIFLOGSLGSSNATLALYALNEHPOTHPITTLQ 122  
 133 DLGLRGLADVPPF 145  
 DB 123 LWARBISDVPPF 135

RESULT 11  
 US-09-634-238-305  
 Sequence 305, Application US/09634238  
 Patent No. 6544772  
 GENERAL INFORMATION:  
 APPLICANT: Glenn, Matthew  
 APPLICANT: Hayakkala, Ilkka J.  
 APPLICANT: Bloksberg, Leonard, N.  
 APPLICANT: Lubbers, Mark W.  
 APPLICANT: Dekker, James  
 APPLICANT: Christensen, Anna C.  
 APPLICANT: Holland, Ross  
 APPLICANT: O'Toole, Paul W.  
 APPLICANT: Reid, Julian R.  
 APPLICANT: Coolbear, Timothy  
 TITLE OF INVENTION: Polynucleotides, materials incorporating  
 TITLE OF INVENTION: them and methods for using them.  
 FILE REFERENCE: 11000.104301  
 CURRENT APPLICATION NUMBER: US/09/634,238  
 CURRENT FILING DATE: 2000-08-08  
 NUMBER OF SEQ ID NOS: 422  
 SOFTWARE: PatsSeq for Windows Version 4.0  
 SEQ ID NO 305  
 LENGTH: 99  
 TYPE: PRT  
 ORGANISM: *Lactobacillus rhamnosus*  
 US-09-634-238-305

Query Match 10.6%; Score 160; DB 4; Length 99;  
 Best Local Similarity 44.4%; Pred. No. 4.9e-10;  
 Matches 36; Conservative 15; Mismatches 30; Indels 0; Gaps 0;

QY 71 DLTVAANLTKSHTCVRCVDCIDIEKNLPMGGGLGGSSDATTLYVLRNLMGLSKRE 130  
 DB 2 NLAAYRAALALQGRANIRGQRHIEHRTIPVAGLGSSDAAAVLGLNLSLQLGYSRQ 61  
 QY 131 LMDLGLRLGADVPVPEVGCSSA 151  
 DB 62 LARIGLQVDSQVPCYSESTA 82

RESULT 12  
 US-08-843-521-2  
 / Sequence 2, Application US/08843521  
 / Patent No. 5753479  
 / GENERAL INFORMATION:  
 / APPLICANT: Lawlor, Elizabeth  
 / TITLE OF INVENTION: No. 5753479el Compounds  
 / NUMBER OF SEQUENCES: 8  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: SmithKline Beecham Corporation  
 / STREET: 709 Swedeland Road  
 / CITY: King of Prussia  
 / STATE: PA  
 / COUNTRY: USA  
 / ZIP: 19406-0939  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Diskette  
 / COMPUTER: IBM Compatible  
 / OPERATING SYSTEM: DOS  
 / SOFTWARE: FASTSEQ for Windows Version 2.0  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/843,521  
 / FILING DATE: 18-APR-1997  
 / CLASSIFICATION: 435  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: 9607993.4  
 / FILING DATE: 18-APR-1996  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Gimm1, Edward R  
 / REGISTRATION NUMBER: 38,891  
 / REFERENCE/DOCKET NUMBER: P31457-6  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 610-270-4478  
 / TELEFAX: 610-270-5090  
 / TELEX:  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 783 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-08-843-521-2

Query Match 6.0%; Score 91.5; DB 1; Length 783;  
 Best Local Similarity 22.4%; Pred. No. 0.38; 76; Indels 39; Gaps 8;  
 Matches 43; Conservative 34; Mismatches 76; Indels 39; Gaps 8;  
 QY 55 GRVTLRNPIS-----GVPEQDILTVRAANLTKSHTCVRCVDCIDIEKNLPMGGGLGGG 107  
 DB 354 GRNLNRSSESSRFKGINVATVNEALDAAASMTAEAGAT-----VRKGISVAGELDTS 407  
 QY 108 SSDAATLTVANRLMGSLSKREIMDLGRLGADVVPVFGCSAMGEGVSEDLQAITLP- 166  
 DB 408 DVEVSSILADVNRKVLGTSLTADYADVPRRLG-----FGLSG-----NADSFIVSVPR 455  
 QY 167 EQWFVILPKDCHVNTGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVYCKLYRFXD- 224  
 DB 456 RRMDDITLADLFEELRIYGYDRLPTS---LPKDDGTAGR-----LTVIQKLRQVRIIA 507  
 QY 225 ---AIDALLCYA 233  
 DB 508 EGAGLTREITTYA 519

QY 55 GRVTLRNPIS-----GVPEQDILTVRAANLTKSHTCVRCVDCIDIEKNLPMGGGLGGG 107  
 DB 354 GRNLNRSSESSRFKGINVATVNEALDAAASMTAEAGAT-----VRKGISVAGELDTS 407  
 QY 108 SSDAATLTVANRLMGSLSKREIMDLGRLGADVVPVFGCSAMGEGVSEDLQAITLP- 166  
 DB 408 DVEVSSILADVNRKVLGTSLTADYADVPRRLG-----FGLSG-----NADSFIVSVPR 455  
 QY 167 EQWFVILPKDCHVNTGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVYCKLYRFXD- 224  
 DB 456 RRMDDITLADLFEELRIYGYDRLPTS---LPKDDGTAGR-----LTVIQKLRQVRIIA 507  
 QY 225 ---AIDALLCYA 233  
 DB 508 EGAGLTREITTYA 519

RESULT 13  
 US-09-012-871-2  
 / Sequence 2, Application US/09012871  
 / Patent No. 6200787  
 / GENERAL INFORMATION:  
 / APPLICANT: Lawlor, Elizabeth  
 / TITLE OF INVENTION: No. 6200787el Compounds  
 / NUMBER OF SEQUENCES: 8  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: SmithKline Beecham Corporation  
 / STREET: 709 Swedeland Road  
 / CITY: King of Prussia  
 / STATE: PA  
 / COUNTRY: USA  
 / ZIP: 19406-0939  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Diskette  
 / COMPUTER: IBM Compatible  
 / OPERATING SYSTEM: DOS  
 / SOFTWARE: FASTSEQ for Windows Version 2.0  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/09/012,871  
 / FILING DATE:  
 / CLASSIFICATION:  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: 08/843,521  
 / FILING DATE:  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Gimm1, Edward R  
 / REGISTRATION NUMBER: 38,891  
 / REFERENCE/DOCKET NUMBER: P31457-6  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 610-270-4478  
 / TELEFAX: 610-270-5090  
 / TELEX:  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 783 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-09-012-871-2

Query Match 6.0%; Score 91.5; DB 3; Length 783;  
 Best Local Similarity 22.4%; Pred. No. 0.38; 76; Indels 39; Gaps 8;  
 Matches 43; Conservative 34; Mismatches 76; Indels 39; Gaps 8;  
 QY 55 GRVTLRNPIS-----GVPEQDILTVRAANLTKSHTCVRCVDCIDIEKNLPMGGGLGGG 107  
 DB 354 GRNLNRSSESSRFKGINVATVNEALDAAASMTAEAGAT-----VRKGISVAGELDTS 407  
 QY 108 SSDAATLTVANRLMGSLSKREIMDLGRLGADVVPVFGCSAMGEGVSEDLQAITLP- 166  
 DB 408 DVEVSSILADVNRKVLGTSLTADYADVPRRLG-----FGLSG-----NADSFIVSVPR 455  
 QY 167 EQWFVILPKDCHVNTGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVYCKLYRFXD- 224  
 DB 456 RRMDDITLADLFEELRIYGYDRLPTS---LPKDDGTAGR-----LTVIQKLRQVRIIA 507  
 QY 225 ---AIDALLCYA 233  
 DB 508 EGAGLTREITTYA 519

QY 55 GRVTLRNPIS-----GVPEQDILTVRAANLTKSHTCVRCVDCIDIEKNLPMGGGLGGG 107  
 DB 354 GRNLNRSSESSRFKGINVATVNEALDAAASMTAEAGAT-----VRKGISVAGELDTS 407  
 QY 108 SSDAATLTVANRLMGSLSKREIMDLGRLGADVVPVFGCSAMGEGVSEDLQAITLP- 166  
 DB 408 DVEVSSILADVNRKVLGTSLTADYADVPRRLG-----FGLSG-----NADSFIVSVPR 455  
 QY 167 EQWFVILPKDCHVNTGEIFSAENLTRNSAVVTMSDFLAGDNNDCEVYCKLYRFXD- 224  
 DB 456 RRMDDITLADLFEELRIYGYDRLPTS---LPKDDGTAGR-----LTVIQKLRQVRIIA 507  
 QY 225 ---AIDALLCYA 233  
 DB 508 EGAGLTREITTYA 519

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
;; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: GTC-007  
;; CURRENT APPLICATION NUMBER: US/09/134,001C  
;; CURRENT FILING DATE: 1998-08-13  
;; PRIOR APPLICATION NUMBER: US 60/064,964  
;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 4141  
;; LENGTH: 315  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4141

Job time : 9.94468 secs

Query Match 5.9%; Score 89; DB 4; Length 315;  
Best Local Similarity 20.0%; Pred. No. 0.19;  
Matches 30; Conservative 38; Mismatches 66; Indels 16; Gaps 5;

QY 63 ISGVPEQDD-LTVRANILKSHTGCVRCVACIDIEKNLPMGGIGGSSDAATLIVLNR 120  
DB 61 LELPFDENNYIYQRLNARKYNTLPSQLIMESDIPLANGSSASALVGLPIANY 120  
QY 121 LMLGLSKRELMDGLRLGA--DPEVFYFG--CSAMGEVSE--DLQAITLPEQMFVI 173  
DB 121 FGNIGSKTEILQLATEIRGHPNVAPITLYGGLIAGFYNPITKITDVARIIEVPHVDILF 180  
QY 174 KPDCHVNT-----GEIISAENLTNSAV 196  
DB 181 IPEYELRTEDSRRLVLPDTFSHKGVQNSAI 210

## RESULT 15

US-09-252-991A-24594  
; Sequence 24594; Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24594  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24594

Query Match 5.7%; Score 86; DB 4; Length 312;  
Best Local Similarity 30.8%; Pred. No. 0.39;  
Matches 44; Conservative 15; Mismatches 58; Indels 26; Gaps 7;

QY 50 HPYDGRVTLRNPIS-GVPEQDDLTVRANILKSHTGCVRCVACIDIEKNLPMGGIGGSS 108  
DB 155 HPGCBGA-----PLSGIPPRILFVTLVLRHPEGMAGVGSAGLDEANGLAGIAL 209  
QY 109 SDAATLIVLNLMLGLSKRELMDGLRLGADVPEVFYFGCSAMGEVSEDLQAITLPEQ 168  
DB 210 QVPEGLVALVLAGVSKRFKRLVGAASGLVEPLFAVLIC-AMLVGLS---ALLP-- 262  
QY 169 W-----FVI---IKDCI 178  
DB 263 WGLAAAGAMLFVVTETILPESH 285

Search completed: January 29, 2004, 15:57:05

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: January 29, 2004, 15:54:19 : Search time 19.4547 Seconds  
(without alignments)  
3044.503 Million cell updates/sec

Title: US-09-941-947a-12  
Perfect score: 1516  
Sequence: 1 MDYAAGGEMRPAPAKNLMLK.....LVFLAKGNGALYKLEEG 285

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB.pep:\*

6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                  | Description         |
|------------|-------|-------------|--------|---------------------|---------------------|
| 1          | 1516  | 100.0       | 285    | US-09-934-903-8     | Sequence 8, Appl 1  |
| 2          | 1516  | 100.0       | 285    | US-09-934-868-74    | Sequence 12, Appl 1 |
| 3          | 1516  | 100.0       | 285    | US-09-941-947a-12   | Sequence 13766, A   |
| 4          | 638   | 42.1        | 279    | US-10-369-493-13766 | Sequence 13766, A   |
| 5          | 620   | 40.9        | 279    | US-10-369-493-301   | Sequence 201, App   |
| 6          | 616   | 40.6        | 279    | US-10-369-493-802   | Sequence 802, App   |
| 7          | 584.5 | 38.6        | 280    | US-10-369-493-21079 | Sequence 21079, A   |
| 8          | 584   | 38.5        | 294    | US-10-369-493-15324 | Sequence 15324, A   |
| 9          | 582   | 38.4        | 276    | US-10-369-493-15690 | Sequence 15690, A   |
| 10         | 582   | 38.4        | 276    | US-10-369-493-16082 | Sequence 16082, A   |
| 11         | 580.5 | 38.3        | 298    | US-10-369-493-17751 | Sequence 17751, A   |
| 12         | 572   | 37.7        | 270    | US-10-369-493-9184  | Sequence 9184, Ap   |
| 13         | 572   | 37.7        | 270    | US-10-369-493-9359  | Sequence 9359, Ap   |
| 14         | 556   | 36.7        | 213    | US-10-369-493-19675 | Sequence 19675, A   |
| 15         | 533   | 35.2        | 233    | US-10-369-493-8419  | Sequence 8419, Ap   |

|    |       |      |     |                     |                    |
|----|-------|------|-----|---------------------|--------------------|
| 16 | 407   | 26.8 | 284 | US-10-369-493-9956  | Sequence 9956, Ap  |
| 17 | 341   | 22.5 | 288 | US-10-369-493-16541 | Sequence 16541, A  |
| 18 | 327   | 21.6 | 289 | US-10-369-493-7851  | Sequence 7851, Ap  |
| 19 | 324   | 21.4 | 289 | US-10-369-493-23046 | Sequence 23046, A  |
| 20 | 320   | 21.1 | 287 | US-10-369-493-9660  | Sequence 9660, Ap  |
| 21 | 317   | 20.9 | 287 | US-10-369-493-17097 | Sequence 17097, A  |
| 22 | 292   | 19.3 | 316 | US-10-369-493-20265 | Sequence 20265, A  |
| 23 | 286.5 | 18.9 | 295 | US-10-369-493-12283 | Sequence 12283, A  |
| 24 | 281   | 18.5 | 297 | US-10-369-493-11123 | Sequence 11123, A  |
| 25 | 278   | 18.3 | 311 | US-10-369-493-1123  | Sequence 1123, A   |
| 26 | 274   | 18.1 | 292 | US-10-369-493-1123  | Sequence 1123, A   |
| 27 | 273.5 | 18.0 | 261 | US-10-369-493-11764 | Sequence 21, App   |
| 28 | 273.5 | 18.0 | 261 | US-10-369-493-14419 | Sequence 14419, A  |
| 29 | 273.5 | 18.0 | 261 | US-10-369-493-14670 | Sequence 14670, A  |
| 30 | 273.5 | 18.0 | 261 | US-10-369-493-15150 | Sequence 15150, A  |
| 31 | 270.5 | 17.8 | 315 | US-10-369-493-2607  | Sequence 2607, Ap  |
| 32 | 267   | 17.6 | 306 | US-09-112-363-186   | Sequence 186, App  |
| 33 | 254   | 16.8 | 240 | US-10-369-493-17829 | Sequence 17829, A  |
| 34 | 250.5 | 16.5 | 268 | US-10-369-493-65    | Sequence 65, Appl  |
| 35 | 239   | 15.8 | 295 | US-10-369-493-16874 | Sequence 16874, A  |
| 36 | 233.5 | 15.4 | 271 | US-10-369-493-3024  | Sequence 3024, Ap  |
| 37 | 216   | 14.2 | 311 | US-09-738-626-4515  | Sequence 4515, Ap  |
| 38 | 203   | 13.4 | 245 | US-10-369-493-10645 | Sequence 10645, A  |
| 39 | 178   | 11.7 | 141 | US-10-369-762-1236  | Sequence 1236, App |
| 40 | 160   | 10.6 | 99  | US-10-264-213-208   | Sequence 208, App  |
| 41 | 127   | 8.4  | 238 | US-09-882-227-358   | Sequence 358, App  |
| 42 | 125   | 8.2  | 314 | US-09-815-242-10963 | Sequence 10963, A  |
| 43 | 116   | 7.7  | 575 | US-10-104-047-2815  | Sequence 2815, Ap  |
| 44 | 116   | 7.7  | 813 | US-10-108-2604-3901 | Sequence 3901, Ap  |
| 45 | 116   | 7.7  | 990 | US-10-094-749-2568  | Sequence 2568, Ap  |

## ALIGNMENTS

RESULT 1

US-09-934-903-8

Sequence 8, Application US/09934903

Patent No. US20020102690A1

GENERAL INFORMATION:

APPLICANT: Koffas, Mattheos

APPLICANT: Odem, J. Martin

APPLICANT: Schenle, Andreas J.

APPLICANT: Tomb, Jean-Francois

APPLICANT: Rouviere, Pierre

APPLICANT: Picataggio, Stephen

APPLICANT: Cheng, Qiong

TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

FILE REFERENCE: C1646 US NA

CURRENT APPLICATION NUMBER: US/09/934,903

CURRENT FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: 60/229,907

PRIOR FILING DATE: September 1, 2001

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Microsoft Office 97

SEQ ID NO 8

LENGTH: 285

TYPE: PRT

ORGANISM: Methylobionas 16a

FEATURE:

OTHER INFORMATION: Amino acid sequences encoded by ORF4

US-09-934-903-8

Query Match 100.0%; Score 1516; DB 10; Length 285;

Best Local Similarity 100.0%; Pred. No. 5.7e-156; Indels 0; Gaps 0;

Matches 285; Conservative 0; Mismatches 0;

CY 1 MDYAAGGEMRPAPAKNLMLRITGRPDYHLLQTVFQWLDLQWLFHPPVDSRYTLR 60

DB 1 MDYAAGGEMRPAPAKNLMLRITGRPDYHLLQTVFQWLDLQWLFHPPVDSRYTLR 60

CY 61 NPISGVEQDDLTFRANLTKSHTCYRGVCIDIEKNLPAGGGIGGSSDAATLVVILNR 120

| Db  | Qy  | Db  | Qy | Db | Qy |
|-----|---|-----|----|----|----|
| 61  | NPISGVPEGGDILTVRAANLTKSHTECVAGVCIIDEIKPLPWGGGLGGSSDAATTLTVLNR | 120 |    |    |    |
| 121 | LMGLGSKREIMDLGLRLGADVPVPLVFGSANGESVSDDAITLPBEMFPIITIPDCHN     | 180 |    |    |    |
| 121 | LMGLGSKREIMDLGLRLGADVPVPLVFGSANGESVSDDAITLPBEMFPIITIPDCHN     | 180 |    |    |    |
| 181 | TGSIIFSAENLTENSAVVTMSDFLAGDNKDCSEVVCCKLYRPVKAIDPALLCYAEARLTGT | 240 |    |    |    |
| 181 | TGSIIFSAENLTENSAVVTMSDFLAGDNKDCSEVVCCKLYRPVKAIDPALLCYAEARLTGT | 240 |    |    |    |
| 241 | GACVFAQPCCKEDAESALGLKDRMTVFLAKLINSALYKYLEGQ                   | 285 |    |    |    |
| 241 | GACVFAQPCCKEDAESALGLKDRMTVFLAKLINSALYKYLEGQ                   | 285 |    |    |    |

RESULT 2  
US-09-934-868-74

Sequence 74, Application US/0934868  
Patent No. US20020137150A1  
GENERAL INFORMATION:  
APPLICANT: Koffas, Matteo  
APPLICANT: Odem, James M  
APPLICANT: Schenzle, Andreas J  
TITLE OF INVENTION: IDENTIFYING METHANOTROPHIC BACTERIAL STRAIN  
FILE REFERENCE: C14596 US NA  
CURRENT APPLICATION NUMBER: US/09/934, 868  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/4229, 858  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 74  
LENGTH: 285  
TYPE: PR1  
ORGANISM: *Methylobomonas 16a*  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ISPE  
US-09-934-868-74

|                           |        |                     |        |                  |
|---------------------------|--------|---------------------|--------|------------------|
| Query Match               | 100.0% | Score 1516;         | DB 10; | Length 285;      |
| Best Local Similarity     | 100.0% | Pred. No. 5.7e-156; |        |                  |
| Matches 285; Conservative | 0;     | Mismatches          | 0;     | Indels 0; Gaps 0 |

|    |     |  |     |
|----|-----|--|-----|
| QY | 1   | MDYAAAGGEMWPAALINMLRITGSRPGVGLIOTVQMDLDDMYLFFHVDGGRVTLR      | 60  |
| Db | 1   | MDYAAAGGEMWPAALINMLRITGSRPGVGLIOTVQMDLDDMYLFFHVDGGRVTLR      | 60  |
| QY | 61  | NPIISGVPEDDDLVVRANLTKSHTCGVGVCIIDEKELPMGGGIGGGSSDAATLVVLNR   | 120 |
| Db | 61  | NPIISGVPEDDDLVVRANLTKSHTCGVGVCIIDEKELPMGGGIGGGSSDAATLVVLNR   | 120 |
| QY | 121 | LMGLGSGKEELMDLGLRGLADVPVVRGCSAMGEGVSESDAATLVPEOMPIVIRPDCHN   | 180 |
| Db | 121 | LMGLGSGKEELMDLGLRGLADVPVVRGCSAMGEGVSESDAATLVPEOMPIVIRPDCHN   | 180 |
| QY | 181 | TGEISPAEMLTRNSAAVVTMSDFLAGDNRRDCEVVCCKYRPVVKOALIDALLCYAARLGT | 240 |
| Db | 181 | TGEISPAEMLTRNSAAVVTMSDFLAGDNRRDCEVVCCKYRPVVKOALIDALLCYAARLGT | 240 |
| QY | 241 | GACVFAQPCNKEDAESALEGLKDRMLVPLAKINOSALYKKELEQG                | 285 |
| Db | 241 | GACVFAQPCNKEDAESALEGLKDRMLVPLAKINOSALYKKELEQG                | 285 |

### RESULT 3

US-09-941-947A-12  
; Sequence 12, Application US/09941947A  
; Publication No. US20030003528A1

```
;; APPLICANT: Brzostowicz, Patricia C.
;; APPLICANT: Cheng, Qiong
;; APPLICANT: DiCosimo, Deana J.
```

```

1 APPLICANT: Koffas, Mattheos
2 APPLICANT: Miller, Edward S. Jr.
3 APPLICANT: Odom, J. Martin
4 APPLICANT: Picataggio, Steve
5 APPLICANT: Rouviere, Pierre E.
6 TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
7 FILE REFERENCE: CL1903 US NA
8 CURRENT APPLICATION NUMBER: US/09/941,947A
9 CURRENT FILING DATE: 2001-09-01
10 PRIOR APPLICATION NUMBER: 60/229,907
11 PRIOR FILING DATE: 2000-09-01
12 PRIOR APPLICATION NUMBER: 60/229,858
13 PRIOR FILING DATE: 2000-09-01
14 NUMBER OF SEQ ID NOS: 60
15 SOFTWARE: Microsoft Office 97
16 SEQ ID NO 12
17 LENGTH: 285
18 TYPE: PRT
19 ORGANISM: Methylobionas 16a
20 US-05-941-947A-12

```

ORGANISM: *Methylobionas* 166  
US-05-941-947A-12

|                       |                |              |          |            |
|-----------------------|----------------|--------------|----------|------------|
| Query Match           | 100.0%         | Score 1516   | DB 11    | Length 285 |
| Best Local Similarity | 100.0%         | Pred. No. 5  | 7e-156   |            |
| Matches 285           | Conservative 0 | Mismatches 0 | Indels 0 | Gaps 0     |

[illegible]

RESULT 4  
US-10-369-493-13766

```

; Sequence 13766, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

```

```

1  APPLICANT: Cao, Yongwei
2  APPLICANT: Hinkle, Gregory J.
3  APPLICANT: Slater, Steven C.
4  APPLICANT: Goldman, Barry S.
5  APPLICANT: Chen, Xianfeng
6  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
7  TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
8  FILE REFERENCE: 38-10(52052)B
9  CURRENT APPLICATION NUMBER: US/10/369,493
10 CURRENT FILING DATE: 2003-02-28
11 PRIOR APPLICATION NUMBER: US 60/360,039
12 PRIOR FILING DATE: 2002-02-21
13 NUMBER OF SEQ ID NOS: 47374
14 SEQ ID NO 13766
15 LENGTH: 279
16
17 TYPE: PRT
18 ORGANISM: Pseudomonas fluorescens
19
20 US-10-369-493-13766

```

|                       |                  |                  |           |             |
|-----------------------|------------------|------------------|-----------|-------------|
| Query Match           | 42.1%;           | Score 638;       | DB 12;    | Length 279; |
| Best Local Similarity | 49.6%;           | Pred. 0.1-2e-60; |           |             |
| Matches 135;          | Conservative 41; | Mismatches 96;   | Indels 0; | Gaps 0;     |

|    | Query Match           | Score 620;   | Length 279;        |
|----|-----------------------|--|--------------------|
|    | Best Local Similarity | 48.0%;   | Pred. No. 1.1e-58; |
|    | Matches 132;          | Conservative 36;   | Mismatches 101;    |
|    |                       | Indels 6;  | Gaps 2;            |
| QY | 11                    | WPAKALNLMKRLTGRPPDGYTHLLQTVQMDLCLWLFPHPPVDGRVTLRNPISGVEED    | 70                 |
| DB | 5                     | WSPKALNLFVLTGRSSDGYHETQLTFQFLDYGDEITTPRQNDQIRLLTPVXGVEND     | 64                 |
| QY | 71                    | DLVFAANLTKSH-----TGC--VRGYCIDIEKNLPMGGGLGGSSDATTLLVNLRLMGL   | 124                |
| DB | 65                    | NLIVRAARLLQHPAEKQSTGTGRGADLHIHRLPLMGSGGLGGSSNATVALIANYMQT    | 124                |
| QY | 125                   | GLSKRELMDGLRLGADVVFVFGCSAMGEGUSELDQALITPEQMFVLIKPDCHVNTGRI   | 184                |
| DB | 125                   | NLSDDLEAQLGVLGADVVFVFKGAAAPFEGIGETKLOPASPEEKVFLAHNGIEITPPKI  | 184                |
| QY | 185                   | FSALNLTNRNSAVVMSDFLAGDNRRDSCSEVCKLTPPKADALDALCYEARLITGTACV   | 244                |
| DB | 185                   | FTDPBELKENSPIRTITPALLQAPFNQDCPIAKRRREVEQGLTSMWLEETPSRLTGTACV | 244                |
| QY | 245                   | PAQFCNKEDAESLGBGLKDRMLVFLAKLINGSALT                          | 279                |
| DB | 245                   | FGFESFSPASARKVLAQAPSMQGFVARGVNISPLH                          | 279                |

RESULT 7  
 US-10-369-493-21079  
 Sequence 21079, Application US/10369493  
 Publication No. US20030233675A1  
 GENERAL INFORMATION:  
 APPLICANT: Cao, Yongwei  
 APPLICANT: Hinkle, Gregory J.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Goldman, Barry S.  
 APPLICANT: Chen, Xianfeng  
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 TIME OF INVENTION:  
 FILE REFERENCE: 38-10(52052) B  
 CURRENT APPLICATION NUMBER: US/10/369,493  
 CURRENT FILING DATE: 2003-02-28  
 PRIOR APPLICATION NUMBER: US 60/360,039  
 PRIOR FILING DATE: 2002-02-21  
 NUMBER OF SEQ ID NOS: 47374  
 SEQ ID NO 21079  
 LENGTH: 280  
 TYPE: PRT  
 ORGANISM: *Xenorhabdus nematophilus*  
 US-10-369-493-21079





Best Local Similarity 46.0%; Pred. No. 1.4e-54;  
Matches 126; Conservative 41; Mismatches 105; Indels 2; Gaps 2;

QY 11 WPAPAKNTLMRTGRPDGTHLLQTVFQMDLDCMLTFHPVDGRV-TLRNPISGVPEQ 69  
DB 3 WPAPAKNTLMRTGRPDGTHLLQTVFQMDLDCMLTFHPVDGRV-TLRNPISGVPEQ 62  
QY 70 DDLTVRAANLTKSHTCVGVICIDIEKNLPMGGGAGGSSDAATTLVYNRLMGLSKR 129  
DB 63 DDLTVRAANLTKSHTCVGVICIDIEKNLPMGGGAGGSSDAATTLVYNRLMGLSKR 122  
QY 130 ELMDLGLRLGADVVFVFGCSAMGSGVSEDLQAITLPEQWFIKPDCHVTGGEIJSAN 189  
DB 123 TLAELGLRLGADVVFVFGCSAMGSGVSEDLQAITLPEQWFIKPDCHVTGGEIJSAN 182  
QY 190 LTRNSAVVWMSDFLAGNDRNCSE-VVCKLYRPVDAIDALCYAARLTGTGACVPAQF 248  
DB 183 LTRNSAVVWMSDFLAGNDRNCSE-VVCKLYRPVDAIDALCYAARLTGTGACVPAQF 242  
QY 249 CNKEDASALEGLKDRMLVFLAKGLNQSALYKXL 282  
DB 243 ATBAAAEQAMPHLPGNLRAWVEGAHSPILDAL 276

RESULT 11  
US-10-369-493-17751  
Sequence 17751, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 17751  
LENGTH: 298  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-17751

Query Match 38.3%; Score 580.5; DB 12; Length 298;  
Best Local Similarity 46.3%; Pred. No. 2.3e-54;  
Matches 132; Conservative 44; Mismatches 106; Indels 3; Gaps 3;

QY 1 MDVAAWGERMPAPAKNTLMRTGRPDGTHLLQTVFQMDLDCMLTFHPVDGRV-TL 59  
DB 11 VDDGVGM-SAWPAPAKNTLMRTGRPDGTHLLQTVFQMDLDCMLTFHPVDGRV-TL 69  
QY 60 RNPISGVPEQDGLTRANLTKSHTCVGVICIDIEKNLPMGGGAGGSSDAATTLVYN 119  
DB 70 RNPISGVPEQDGLTRANLTKSHTCVGVICIDIEKNLPMGGGAGGSSDAATTLVYN 129  
QY 120 RUMGLSLKSLMDLGLRLGADVVFVFGCSAMGSGVSEDLQAITLPEQWFIKPDCHV 179  
DB 130 RUMGLSLKSLMDLGLRLGADVVFVFGCSAMGSGVSEDLQAITLPEQWFIKPDCHV 189  
QY 180 NTGEIJSANLTRNSAVVWMSDFLAGNDRNCSE-VVCKLYRPVDAIDALCYAARLT 238  
DB 190 NTGEIJSANLTRNSAVVWMSDFLAGNDRNCSE-VVCKLYRPVDAIDALCYAARLT 249  
QY 239 GTGACVPAQFCNKEDASALEGLKDRMLVFLAKGLNQSALYKXL 283  
DB 250 GTGACVPAQFCNKEDASALEGLKDRMLVFLAKGLNQSALYKXL 294

RESULT 12  
US-10-369-493-9184

Sequence 9184, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 9184  
LENGTH: 270  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-9184

Query Match 37.7%; Score 572; DB 12; Length 270;  
Best Local Similarity 47.4%; Pred. No. 1.7e-53;  
Matches 128; Conservative 42; Mismatches 98; Indels 2; Gaps 2;

QY 11 WPAPAKNTLMRTGRPDGTHLLQTVFQMDLDCMLTFHPVDGRV-TLRNPISGVPEQ 69  
DB 1 WPAPAKNTLMRTGRPDGTHLLQTVFQMDLDCMLTFHPVDGRV-TLRNPISGVPEQ 60  
QY 70 DDLTVRAANLTKSHTCVGVICIDIEKNLPMGGGAGGSSDAATTLVYNRLMGLSKR 129  
DB 61 DDLTVRAANLTKSHTCVGVICIDIEKNLPMGGGAGGSSDAATTLVYNRLMGLSKR 120  
QY 130 ELMDLGLRLGADVVFVFGCSAMGSGVSEDLQAITLPEQWFIKPDCHVTGGEIJSAN 189  
DB 121 ELMDLGLRLGADVVFVFGCSAMGSGVSEDLQAITLPEQWFIKPDCHVTGGEIJSAN 180  
QY 190 LTRNSAVVWMSDFLAGNDRNCSE-VVCKLYRPVDAIDALCYAARLTGTGACVPAQF 248  
DB 181 LTRNSAVVWMSDFLAGNDRNCSE-VVCKLYRPVDAIDALCYAARLTGTGACVPAQF 240  
QY 249 CNKEDASALEGLKDRMLVFLAKGLNQSALYKXL 278  
DB 241 STRDEACALERLPGNLRAWVEGAHSPILDAL 270

RESULT 13  
US-10-369-493-9359

Sequence 9359, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xiandeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 9359  
LENGTH: 270  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-9359

Query Match 37.7%; Score 572; DB 12; Length 270;  
 Best Local Similarity 47.4%; Pred. No. 1.7e-53;  
 Matches 128; Conservative 42; Mismatches 98; Indels 2; Gaps 2;

QY 11 WPAPAKNLMLRITGRPPDGYHLQTVFQMLDLCMDLTFHPVDDGRVTLNRPISGVEEQ 69  
 DB 1 WPAPAKNLMLFQITGRVVGYHQLQTVFRLDMGDTIHLRVBDDQIHRIGSVTGVVEA 60  
 QY 70 DDLTAAANLKSHTGCVRGVCIDIEKRLPQGGGCGSSDAATLVYANLMLGSLSKR 129  
 DB 61 DDLVRAAALLHARNVYIGADIPEKRIEVGSGGSSDAATLVYANLMLHTRLDVA 120  
 QY 130 ELMDLGLRLGADVPPVFGSCAMGCVSEDLQATITLPEQMFYIKPDCHVNTGEIFSAEN 189  
 DB 121 VLAAGLRLGADVPPVFGSCAMGCVSEDLQATITLPEQMFYIKPDCHVNTGEIFSAEN 180  
 QY 190 LTRNSAVVTMSDFLAGDNNDCEB-VCKLTPYKDAIDALCYAERLGTGACVFAOF 248  
 DB 181 LTRDASPATIGDFIAGTAFGNAPFBLRRESAVAGALDMLSRVGFATVTSGCCFVEF 240  
 QY 249 CNKEDASALEGLKDRWLVLAKGNOSAL 278  
 DB 241 STRDEAECHLERLPYGLCANVADGASRSL 270

RESULT 14  
 US-10-369-493-19675  
 ; Sequence 19675, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 19675  
 ; LENGTH: 213  
 ; TYPE: PRT  
 ; ORGANISM: Nitrosomonas europaea  
 ; US-10-369-493-19675

Query Match 36.7%; Score 556; DB 12; Length 213;  
 Best Local Similarity 51.6%; Pred. No. 6.6e-52;  
 Matches 110; Conservative 34; Mismatches 67; Indels 2; Gaps 2;

QY 11 WPAPAKNLMLRITGRPPDGYHLQTVFQMLDLCMDLTFHPVDDGRVTLNRPISGVEEQ 70  
 DB 1 WPAPAKNLMLFQITGRVVGYHQLQTVFRLDMGDTIHLRVBDDQIHRIGSVTGVVEA 60  
 QY 71 DDLTAAANLKSHTGCVRGVCIDIEKRLPQGGGCGSSDAATLVYANLMLGSLSKR 129  
 DB 61 DDLVRAAALLHARNVYIGADIPEKRIEVGSGGSSDAATLVYANLMLHTRLDVA 120  
 QY 130 ELMDLGLRLGADVPPVFGSCAMGCVSEDLQATITLPEQMFYIKPDCHVNTGEIFSAEN 189  
 DB 121 VLAAGLRLGADVPPVFGSCAMGCVSEDLQATITLPEQMFYIKPDCHVNTGEIFSAEN 180  
 QY 190 LTRNSAVVTMSDFLAGDNNDCEB-VCKLTPYKDAIDALCYAERLGTGACVFAOF 248  
 DB 181 LTRDASPATIGDFIAGTAFGNAPFBLRRESAVAGALDMLSRVGFATVTSGCCFVEF 240

RESULT 15  
 US-10-369-493-8419  
 ; Sequence 8419, Application US/10369493

Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 8419  
 ; LENGTH: 233  
 ; TYPE: PRT  
 ; ORGANISM: Ralstonia metallidurans  
 ; US-10-369-493-8419

Query Match 35.2%; Score 533; DB 12; Length 233;  
 Best Local Similarity 49.1%; Pred. No. 2.4e-49;  
 Matches 114; Conservative 35; Mismatches 77; Indels 6; Gaps 2;

QY 53 DGGVTLNRPISGVPEODDLTVRAANLKSHTGCVRGVCIDIEKRLPQGGGCGSSDA 112  
 DB 2 DDGTVIRVFTDIPGPADTDLVRAAARMAQASGVFCVDIADIKILPMGGGIGGSSDA 61  
 QY 113 TTLVYANLMLGSLSKRELMDLGLRLGADVPPVFGSCAMGCVSEDLQATITLPEQMFV 172  
 DB 62 TTLVYANLMLGSLSKRELMDLGLRLGADVPPVFGSCAMGCVSEDLQATITLPEQMFV 121  
 QY 173 IKPDCHVNTGEIFSAENLTRNSAVVTMSDFLAGDN-----RNDCEVVCCKLYRSPKDAID 227  
 DB 122 IHPQGHVPTALISDELTNSPISIVADPACCTNFAFRNDLETATAKFGVARALE 181  
 QY 228 ALICYA-EARLTGACVFAOFCKEDASALGLDRMLVFLAKGNOSAL 278  
 DB 182 WLKNSPHAMWTGSGACVFAFDEDEQTVQRMERLSEWDGRCVKSLSHHPL 233

Search completed: January 29, 2004, 16:21:17  
 Job time : 20.4547 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:44:49 ; Search time 8.38564 Seconds  
(without alignments)  
3268.453 Million cell updates/sec

Title: US-09-941-947a-12

Perfect score: 1516

Sequence: 1 MDYAGNGERMPAPAKNLNM.....LVPLAKINGSLYKLEEG 285

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description                          |
|------------|-------|-------------|--------|----------|--------------------------------------|
| 1          | 668   | 44.1        | 289    | 2 F83063 | isopentenyl monophosphate 4-diphosph |
| 2          | 652.5 | 43.0        | 282    | 2 AH0245 | probable 4-diphosph                  |
| 3          | 629   | 41.5        | 283    | 2 S27732 | conserved hypotet                    |
| 4          | 629   | 41.5        | 283    | 2 AE0720 | isopentenyl monoph                   |
| 5          | 624.5 | 41.2        | 281    | 2 A81875 | hypothetical prote                   |
| 6          | 616   | 40.6        | 283    | 2 A90843 | ychb protein (simi                   |
| 7          | 616   | 40.6        | 283    | 2 B47706 | kinase, GMP fami                     |
| 8          | 616   | 40.6        | 295    | 2 B82109 | conserved hypotet                    |
| 9          | 615   | 40.6        | 281    | 2 B81149 | hypothetical prote                   |
| 10         | 609   | 40.2        | 283    | 2 P85700 | conserved hypotet                    |
| 11         | 606.5 | 40.0        | 321    | 2 A64171 | isopentenyl monoph                   |
| 12         | 580.5 | 38.3        | 298    | 2 P82532 | conserved hypotet                    |
| 13         | 551   | 36.3        | 221    | 2 S49374 | hypothetical prote                   |
| 14         | 493.5 | 32.6        | 294    | 2 G84949 | hypothetical prote                   |
| 15         | 327.5 | 21.6        | 299    | 2 A82654 | hypothetical prote                   |
| 16         | 327.5 | 21.6        | 299    | 2 F97435 | conserved hypotet                    |
| 17         | 324   | 21.3        | 289    | 2 S66075 | B. subtilis YabH p                   |
| 18         | 323   | 21.3        | 291    | 2 A71461 | 4-diphosphocytidyl                   |
| 19         | 323.5 | 21.3        | 293    | 2 AC3444 | B. subtilis YabH p                   |
| 20         | 320   | 21.1        | 299    | 2 AG1098 | B. subtilis YabH p                   |
| 21         | 317   | 20.9        | 287    | 2 E83657 | hypothetical prote                   |
| 22         | 303   | 20.0        | 280    | 2 H89815 | isopentenyl monoph                   |
| 23         | 301.5 | 19.9        | 282    | 2 A97257 | ripening-associate                   |
| 24         | 277   | 18.3        | 401    | 2 T07419 | 4-diphosphocytidyl                   |
| 25         | 275   | 18.1        | 383    | 2 T02642 | ripening-associate                   |
| 26         | 272   | 17.9        | 317    | 2 AC2209 | 4-diphosphocytidyl                   |
| 27         | 270.5 | 17.8        | 315    | 2 S74513 | hypothetical prote                   |
| 28         | 267   | 17.6        | 306    | 2 F70603 | hypothetical prote                   |
| 29         | 260   | 17.2        | 299    | 2 E75254 | conserved hypotet                    |

|    |       |      |     |          |                     |
|----|-------|------|-----|----------|---------------------|
| 30 | 256.5 | 16.9 | 283 | 2 G81732 | kinase, GMP fami    |
| 31 | 255.5 | 16.9 | 291 | 2 B71333 | conserved hypotet   |
| 32 | 254   | 16.8 | 311 | 2 B86939 | probable isopenten  |
| 33 | 250.5 | 16.5 | 268 | 2 A70379 | conserved hypotet   |
| 34 | 239   | 15.8 | 295 | 2 A87415 | kinase, GMP fami    |
| 35 | 236   | 15.6 | 288 | 2 B71470 | probable kinase -   |
| 36 | 233.5 | 15.4 | 271 | 2 H72261 | conserved hypotet   |
| 37 | 178   | 11.7 | 141 | 2 H86609 | kinase (imported)   |
| 38 | 167   | 11.0 | 255 | 2 C81314 | hypothetical prote  |
| 39 | 145.5 | 9.6  | 274 | 2 F71820 | hypothetical prote  |
| 40 | 134   | 8.8  | 265 | 2 D82871 | conserved hypotet   |
| 41 | 132.5 | 8.7  | 322 | 2 F82583 | homoserine kinase   |
| 42 | 132   | 8.7  | 268 | 2 C64700 | conserved hypotet   |
| 43 | 125   | 8.2  | 314 | 2 F64047 | homoserine kinase   |
| 44 | 119   | 7.8  | 309 | 2 F84952 | homoserine kinase   |
| 45 | 116   | 7.7  | 990 | 2 J07878 | lucokinaase (EC 2.7 |

## ALIGNMENTS

## RESULT 1

F83063 isopentenyl monophosphate kinase PA4669 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: F83063

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bri-

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,

Loay, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathoc

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: F83063

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-282 <STO>

A:Cross-references: GB:AE004880; GB:AE004091; NID:G9950932; PIDN:AA08056.1; GSPDB:GN001;

A:Experimental source: strain PA01

C:Genetics:

A:Gene: ipk; PA4669

C:Superfamily: conserved hypothetical protein HP1443

Query Match 44.1%; Score 668; DB 2; Length 282;  
Best Local Similarity 51.8%; Pred. No. 1.1e-52;  
Matches 141; Conservative 36; Mismatches 95; Indels 0; Gaps 0;

|    |     |   |     |
|----|-----|---|-----|
| QY | 12  | PAPAKNMLRIRGRPDGTHILQTPQMDLGDWLFTHFVDDGRVTLAMPISGPEQDD    | 71  |
| DB | 8   | PAPAKNMLRIRGRPDGTHILQTPQMDLGDWLFTHFVDDGRVTLAMPISGPEQDD    | 67  |
| QY | 72  | LTVPAANTLAKSHGCVRGVCIDIEKRLPMGGGLGGSSDAATTLVNLRLMGLSLREL  | 131 |
| DB | 68  | LTVPAANTLAKSHGCVRGVCIDIEKRLPMGGGLGGSSDAATTLVNLRLMGLSLREL  | 127 |
| QY | 132 | MDGLRLGAVPVFVPGCSAMGCVSEDLQATLPQMTPIITPDGCHVNGEIPSAEMLT   | 191 |
| DB | 128 | MDGLRLGAVPVFVPGCSAMGCVSEDLQATLPQMTPIITPDGCHVNGEIPSAEMLT   | 187 |
| QY | 192 | RNSAVVMTSDPLAGDNNDSEVVCILYRPVKAIDALLCYAARLTGTGACVFAQPCNK  | 251 |
| DB | 188 | RDSALIKVTRVLDGSDNDQPFVBERRYEVRNALILNKFASARLTGTGACVFAQPCNK | 247 |
| QY | 252 | EDASALIEGLKDRMLVPLAKINGSLYKLE                             | 283 |
| DB | 248 | AEADKVSALLPDLQRFVAAKSNISMLHRLS                            | 279 |

## RESULT 2

AH0245 probable 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.-) [imported] - Yers

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001



A:Residues: 1-281 <PAR>  
 A:Cross-references: GB:AL162755; GB:AL157959; NID:97379742; PIDN:QAB84355.1; PID:9737978  
 A:Experimental source: serogroup A, strain Z2491  
 C:Genetics:  
 A:Gene: NKA1092  
 C:Superfamily: conserved hypothetical protein HP1443

Query Match 41.2%; Score 624.5; DB 2; Length 281;  
 Best Local Similarity 46.7%; Pred. No. 9,3e-49;  
 Matches 128; Conservative 51; Mismatches 88; Indels 7; Gaps 2;

QY 9 ERMPPAKNLMRLRTGRPDGYHLLQTVFQMLDCMILTFHVDGRTLNPIGSGVE 68  
 DB 8 QAFPAKNTLDRITGRREDGYHNIIESIFCLIDQLQTVLKRDDDKITLHPVDMQ 67  
 QY 69 QDDLTVRAANLLKSHGCVAGVCIDIEKNLPMGGGIGGSSDAATLVNLMGLSK 128  
 DB 68 EADLSTRMSSLKRYARTPTGVEIWMDDKIPYAGIGGSSDAATLVNLMWQGLTQ 127  
 QY 129 RELMDLGLRLGADVPVFGCSAMGCVSEDLQATLPEQWVILIKPDCHVTGIFSAE 188  
 DB 128 RQIDSGALGADVPFIFGKNAPAGIGRLEMDIPQWVITVDPVHVSIAKIFTHS 187  
 QY 189 NLRNAGVVTMSOFLAGDN---RNDCEVYCKLTPYVDALDALLCYABRLTGACV 244  
 DB 188 GLTRNSASSIMPTF---QMQPRNDWQAVFVEYEVKXVSLRYPALMTSGACV 244  
 QY 245 FAQCNKEDAESALEGLKDRMLVFAKGNQSL 278  
 DB 245 FTACQDRNSAVNITROYSDLYEAYLAEKSKHL 278

## RESULT 6

YchB protein [similarity] - Escherichia coli (strain O157:H7, substrain RMD 0509952)  
 A:Accession: A90843  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence, revision 18-Jul-2001 #ext, change 02-Nov-2001  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gaswara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: A90843  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-283 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA035136.1; PID:913361178; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 A:Gene: Ecs1713  
 C:Superfamily: conserved hypothetical protein HP1443

Query Match 40.6%; Score 616; DB 2; Length 283;  
 Best Local Similarity 47.9%; Pred. No. 5.5e-48;  
 Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;

QY 10 RWPAPKNTLMRLRTGRPDGYHLLQTVFQMLDCMILTFHVDGRTLNPIGSGVEQ 69  
 DB 4 QWSPAKNTLMFLYITGGRADGYHLLQTVFQMLDCMILTFHVDGRTLNPIGSGVEH 63  
 QY 70 DDLTVRAANLL-----KSHGCVAGVCIDIEKNLPMGGGIGGSSDAATLVNLM 121  
 DB 64 DNLIVRAARLAKTAAADSGRLPTG--SGANISIDKRLPMGGGIGGSSNAATLVNLM 121  
 QY 122 WGLGSKRELMDLGLRLGADVPVFGCSAMGCVSEDLQATLPEQWVILIKPDCHVT 181  
 DB 122 WQCGISMDLAEGLTLAGADVPVFGHAAFAEGVEILTPVDPKMYLVAFHGSVSIPT 181  
 QY 182 GEIPEAKNTLRTGRPDGYHLLQTVFQMLDCMILTFHVDGRTLNPIGSGVEQ 241  
 DB 182 PVLTFDPELPRNTPKRSITETLKCEFSNDCEVIAKRRFSVDAVLSWLEIAPSRITG 241

QY 242 ACVPAQNKDASALEGLKDRMLVFAKGNQSLYKTL 282  
 DB 242 ACVPAQNKDASALEGLKDRMLVFAKGNQSLYKTL 282

## RESULT 7

YchB protein - Escherichia coli (strain K-12)  
 A:Accession: B47706  
 C:Date: 19-Dec-1993 #sequence, revision 18-Nov-1994 #ext, change 01-Mar-2002  
 C:Accession: B47706; 183566; E64867; P61178; S27569  
 R:Post, D.A.; Hove-Jensen, B.; Siltzer, R.L.  
 J. Gen. Microbiol. 139, 259-266, 1993  
 A:Title: Characterization of the hema-prs region of the Escherichia coli and Salmonella  
 A:Reference number: B47706; MUID:93171869; PMID:7679716  
 A:Accession: B47706  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-283 <POS>  
 A:Cross-references: EMBL:M77237; NID:9147380; PIDN:AA24434.1; PID:9147382  
 A:Experimental source: strain K-12  
 A>Note: sequence extracted from NCBI backbone (NCBI:125627, NCBI:125631)  
 R:Strømmer, H.; Remler, P.; Renner, W.; Hogenauer, G.  
 J. Bacteriol. 177, 4468-4500, 1995  
 A:Title: Expression of genes *kdsA* and *kdsB* involved in 3-deoxy-D-manno-octulosonic acid  
 A:Reference number: I60364; MUID:95362678; PMID:7543480  
 A:Accession: I60364  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-283 <RES>

A:Cross-references: EMBL:U18555; NID:9968925; PIDN:AA64343.1; PID:9968927

A:Experimental source: strain K-12

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: E64867

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-283 <BLAT>

A:Cross-references: GB:AE000219; GB:U00096; NID:91787453; PIDN:AA674292.1; PID:91787459;

A:Experimental source: strain K-12, substrain MG1655

R:Ikemura, M.; Murakami, K.; Hashimoto, M.; Murooka, Y.

Gene 121, 127-132, 1992

A:Title: Cloning and characterization of genes involved in the biosynthesis of delta-amit

A:Reference number: J01361; MUID:93051347; PMID:1427085

A:Accession: PC1178

A:Molecule type: DNA

A:Residues: 1-283 <HAY>

A:Cross-references: DDBJ:D10264; NID:9216522; PIDN:BA00106.1; PID:9216523

A:Genetics:

A:Gene: ychB

A:Map position: 26 min

C:Superfamily: conserved hypothetical protein HP1443

Query Match 40.6%; Score 616; DB 2; Length 283;  
 Best Local Similarity 47.9%; Pred. No. 5.5e-48;  
 Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;

QY 10 RWPAPKNTLMRLRTGRPDGYHLLQTVFQMLDCMILTFHVDGRTLNPIGSGVEQ 69  
 DB 4 QWSPAKNTLMFLYITGGRADGYHLLQTVFQMLDCMILTFHVDGRTLNPIGSGVEH 63  
 QY 70 DDLTVRAANLL-----KSHGCVAGVCIDIEKNLPMGGGIGGSSDAATLVNLM 121  
 DB 64 DNLIVRAARLAKTAAADSGRLPTG--SGANISIDKRLPMGGGIGGSSNAATLVNLM 121  
 QY 122 WGLGSKRELMDLGLRLGADVPVFGCSAMGCVSEDLQATLPEQWVILIKPDCHVT 181  
 DB 122 WQCGISMDLAEGLTLAGADVPVFGHAAFAEGVEILTPVDPKMYLVAFHGSVSIPT 181  
 QY 182 GEIPEAKNTLRTGRPDGYHLLQTVFQMLDCMILTFHVDGRTLNPIGSGVEQ 241  
 DB 182 PVLTFDPELPRNTPKRSITETLKCEFSNDCEVIAKRRFSVDAVLSWLEIAPSRITG 241

QY 182 GFISAEINTNSAVVMSDFLAGDNNDCEVCKLYRPVKDAIDALCYAARLTGTG 241  
 DB 182 PVIKPDPELRPTPKRSITFLTKCFESNCEVIARRRFEVAVLSWLEFVAPSLTGTG 241  
 QY 242 ACVFAQFCCKEDAESLSEGLKDRML-VFLAKGINOSALYKYL 282  
 DB 242 ACVFAEFDETESEARQVLEQAPF-MELNGFAKAGANLSPILHRAM 282

## RESULT 8

Query Match 40.6%; Score 616; DB 2; Length 295;  
 Best local similarity 47.1%; Pred. No. 5, 8e-48;  
 Matches 128; Conservative 46; Mismatches 98; Indels 0; Gaps 0;  
 C:Species: Vibrio cholerae  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: E82109  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Drygo, I.; Sellers, F.  
 L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: E82109  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-295 <HEI>  
 A:Cross-references: GB:AE004289; GB:AE003852; NID:99656725; PIDN:AAF95327.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC2182  
 A:Map position: 1  
 C:Superfamily: conserved hypothetical protein HP1443

QY 11 WPAAPAKLMLRTTGRPDGYHLIQVFGMLCDMLTFHPVDDGRVTLNPISGVPEQ 70  
 DB 13 WSPAPALNPLVITGRRAGYHDLQTLFQHLGDELTITANNSGNTLSPALVALED 72  
 QY 71 DLTVRAANLKSHTGCVRGVCIIEKNLPMGGGLGGSSDAATTLVNLRLMGLASRE 130  
 DB 73 NLIYKAAVMAKXNAAGSPICADILQHKVLPMMGGIGGGSSNATITLVNLYMLQSTLSDQ 132  
 QY 131 LMDLGRRLADVPVYFGCSAMGEGVSEDLQAITLPEQFVYIKDCVNTGEISAEHL 190  
 DB 133 LAEIGALADVPVYFGFMAAABGVBELSAVBEEKVYLVRAVVSATDITFHPQL 192  
 QY 191 TNSAVVTVSDFLAGDNNDCEVCKLYRPVKDAIDALCYAARLTGTGACVAPFCN 250  
 DB 193 MENTPRDLASLITTPYENDCEKIVRSLYPRVDKQSLMLQVAPSRITQTSQVAPFSS 252  
 QY 251 KEDASALEGLKDRMLVFLAKGINOSALYKYL 282  
 DB 253 RQDAQVFAQLSDNVLAFAVAGRNVSPLKXTL 284

## RESULT 9

QY 182 GFISAEINTNSAVVMSDFLAGDNNDCEVCKLYRPVKDAIDALCYAARLTGTG 241  
 DB 182 PVIKPDPELRPTPKRSITFLTKCFESNCEVIARRRFEVAVLSWLEFVAPSLTGTG 241  
 QY 242 ACVFAQFCCKEDAESLSEGLKDRML-VFLAKGINOSALYKYL 282  
 DB 242 ACVFAEFDETESEARQVLEQAPF-MELNGFAKAGANLSPILHRAM 282

A:Residues: 1-281 <TEI>  
 A:Cross-references: GB:AE002439; GB:AE002098; NID:97226100; PIDN:AAF41285.1; PID:9722611  
 A:Experimental source: serogroup B, strain MCS8  
 C:Genetics:  
 A:Gene: NMB0874  
 C:Superfamily: conserved hypothetical protein HP1443

Query Match 40.6%; Score 615; DB 2; Length 281;  
 Best local similarity 46.1%; Pred. No. 6, 7e-48;  
 Matches 130; Conservative 53; Mismatches 91; Indels 8; Gaps 3;

QY 1 MDVAGWGERPAPAKLMLRTTGRPDGYHLIQVFGMLCDMLTFHPVDDGRVTLR 60  
 DB 1 MNIADG-RQAPSAPAKLMLRTTGRPDGYHLIQVFGMLCDMLTFHPVDDGRVTLR 59  
 QY 61 NPIGVPQDDLTVRANLKSHTGCVRGVCIIEKNLPMGGGLGGSSDAATTLVNLRL 120  
 DB 60 NPVGMFQEVDSFASALQKXANPAGVETLWDDKLPFGAGLGGSSDAATTLVNLRL 119  
 QY 121 LMGGLSKRELMDGLRLGADVVPVFGCSAMGEGVSEDLQAITLPEQFVYIKDCVHN 180  
 DB 120 MWQGLRQRLIDGALLGADVPEFFIRKNAFARGIGBRLDEKMDIPQWVYVAPVHVS 179  
 QY 181 TGEIFSAEINTNSAVVMSDFLAGDN----RNDCEVCKLYRPVKDAIDALCYAAR 236  
 DB 180 TAKIFTHESLITNSASSIMPTF--QNLQPFPRMDQAVFEKYPEVWKAVERSELRGPA 236  
 QY 237 LTGTGACVFAQFCCKEDAESLSEGLKDRMLVFLAKGINOSAL 278  
 DB 237 MTSGGACVFTACQDRNANVYRQVSDLYEAYLAEGSLKPHL 278

## RESULT 10

Query Match 40.2%; Score 609; DB 2; Length 283;  
 Best local similarity 47.5%; Pred. No. 2, 4e-47;  
 Matches 134; Conservative 39; Mismatches 97; Indels 12; Gaps 4;  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: F85700  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: F85700  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-283 <STO>  
 A:Cross-references: GB:AE005174; NID:912514921; PIDN:AAG56066.1; GSPDB:GN00145; UMG:219  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: ycbB  
 C:Superfamily: conserved hypothetical protein HP1443

QY 10 RWPAPAKLMLRTTGRPDGYHLIQVFGMLCDMLTFHPVDDGRVTLNPISGVPEQ 69  
 DB 4 QWSPAPALNPLVITGRRAGYHDLQTLFQHLGDELTITANNSGNTLSPALVALED 72  
 QY 70 DLTVRAANLKSHTGCVRGVCIIEKNLPMGGGLGGSSDAATTLVNLRL 121  
 DB 64 DNLIYKAAVMAKXNAAGSPICADILQHKVLPMMGGIGGGSSNATITLVNLYMLQSTLSDQ 121  
 QY 122 WGLGLSKRELMDGLRLGADVVPVFGCSAMGEGVSEDLQAITLPEQFVYIKDCVHN 181  
 DB 122 WQGLSLMDLAEGLTLAGADVVPVFGFMAAABGVBELSAVBEEKVYLVRAVVSATDITFHPQL 181  
 QY 182 GFISAEINTNSAVVMSDFLAGDNNDCEVCKLYRPVKDAIDALCYAARLTGTG 241  
 DB 182 PVIKPDPELRPTPKRSITFLTKCFESNCEVIARRRFEVAVLSWLEFVAPSLTGTG 241





QY 192 RNSAVTMDFLAGNRNDCEVCKLTPPVKDA 225  
 DB 188 RDSFPAIKVTVLEGDSEKNDQCPVERRRYPBVRNA 221

## RESULT 14

G84949  
 hypothetical protein ychB [imported] - Buchnera sp. (strain APS)

C/Species: Buchnera sp.

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C/Accession: G84949

R/Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A/Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A

A/Reference number: A84930; NCBI:20445173; PMID:10993077

A/Accession: G84949

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-294 <STO>

A/Cross-references: GB:AP000398; GSPDB:GN00144

A/Experimental source: strain APS

A/Genetics:

A/Gene: ychB; B0170

C/Superfamily: conserved hypothetical protein HP1443

Query Match 32.6%; Score 493.5; DB 2; Length 294;  
 Best Local Similarity 40.6%; Pred. No. 6,7e-37;  
 Matches 113; Conservative 48; Mismatches 108; Indels 9; Gaps 2;

QY 13 APAKINLMRTGRPRPGYHLLQTFQMLDCLMTLTFHPVDGRVTLRNPISGVBDL 72

DB 15 SPAKINLFLYVTRGRKXGYNHNIQTLEQFLDYGDFKLIANKGNIELFTEKKI FNNVQNS 74

QY 73 TVRAANLTKSHGCVR-----GVCIDIKNLPMGGGLGGSSDAATTVLVRNL 124

DB 75 IITAAKLKXK-TALLGCKLQNSYGAKITFLKKNIPMGSGLGSSSRATTVLVRNLMT 133

QY 125 GLSKRELMDLGLRGAADVFPVFGCSANGSGVSDLOAITLPQWFIIPKDCNVNTGEI 184

DB 134 QVTLKELSLGLRIGADVGFVWNTAVLEGIDILYPIVQKEKMYLVVYPCINISTRYM 193

QY 185 FSAENITRNSAVTMDFLAGNRNDCEVCKLTPPVKDAIDALICVAEARLTGTGACV 244

DB 194 FSSPFLMSNTAKSLQVLTLPKDFENIARKQFPIKLLIRMLSSYAPSRMTGTGSCV 253

QY 245 PAQCNKEDASALBGLKDRWVFLAKGLNQSALYKCL 282

DB 254 FSEFDNKGAKQKIFSVLPKRVGFIKSNVISPLHNTL 291

QY 254 FSEFDNKGAKQKIFSVLPKRVGFIKSNVISPLHNTL 291

## RESULT 15

AB2654  
 hypothetical protein ipk [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C/Accession: AB2654

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erege, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavina, T.; Levy, R.; Li, M.; McClellan

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, B.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; NCBI:21608550; PMID:11743193

A/Accession: AB2654

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-299 <KUR>

A/Cross-references: GB:AB008688; PDB:1AL41648.1; PDB:1G1738989; GSPDB:GN00186

A/Experimental source: strain C58 (Dupont)

A/Genetics:

A/Gene: ipk

A/Map position: circular chromosome

C/Superfamily: conserved hypothetical protein HP1443

Query Match 21.6%; Score 327.5; DB 2; Length 299;

Best Local Similarity 35.6%; Pred. No. 6.8e-22;

Matches 99; Conservative 41; Mismatches 111; Indels 27; Gaps 10;

QY 13 APAKINLMRTGRPRPGYHLLQTFQMLDCLMTLTFHPVDGRVTLRNPISGVBDL 68

DB 16 APAKINLALHTVTRGRPRPGYHLLQTFQMLDCLMTLTFHPVDGRVTLRNPISGVBDL 75

QY 69 QDDLTVRANLTK--SHTG-CVRGVCIDIKNLPMGGGLGGSSDAATTVLVRNLMT 124

DB 76 GDNLTVRARLDLRALASTGQPARPVHIEKNLPVASGIGGSADAATLRGLRHWDA 135

QY 125 GLSKRELMDLGLRGAADVFPVFGCSANGSGVSDLOAITLPQWFIIPKDCNVNTGEI 183

DB 136 AIAPEKKSIALKLGADVPMCLASRPLIARGIGDIEALTDLPILSWVLNPLAVSTPE 195

QY 184 IFSANITRNSAVTMS-----DPLAGNRNDCEVCKLTPPVKDAIDALICVA 233

DB 196 IFRRLQNRVNPFLTPSTIGATTGMDFLA-QSRNDLQPPAALL-PEIGELTGLSEE 253

QY 234 EA---RLTGTA-C-VPAQCNKEDASALBGLKDRW 265

DB 254 GATLVFMGSGATCFGIFHSFDAKNAETSLKRRPGW 291

Search completed: January 29, 2004, 15:55:31

UDB time : 9.38564 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:41:08 ; Search time 5.36661 Seconds  
(without alignments)  
2497.314 Million cell updates/sec

Title: US-09-941-947A-12  
Perfect score: 1516  
Sequence: 1 MRYAGMGRWRPAPAKNLNM.....LVFLAKGLNQSLYKKEEG 285

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 668   | 44.1        | 282    | 1  | ISPE_PSEAE  |
| 2          | 665   | 43.9        | 289    | 1  | ISPE_RALSO  |
| 3          | 652.5 | 43.0        | 299    | 1  | ISPE_YARPE  |
| 4          | 644.5 | 42.5        | 284    | 1  | ISPE_SHEON  |
| 5          | 642   | 42.3        | 290    | 1  | ISPE_VIBVU  |
| 6          | 636   | 42.0        | 295    | 1  | ISPE_VIBPA  |
| 7          | 629   | 41.5        | 283    | 1  | ISPE_SALTI  |
| 8          | 629   | 41.5        | 283    | 1  | ISPE_SALTY  |
| 9          | 624.5 | 41.2        | 281    | 1  | ISPE_NEIMA  |
| 10         | 624   | 40.9        | 295    | 1  | ISPE_PASMU  |
| 11         | 620   | 40.9        | 283    | 1  | ISPE_ECOL6  |
| 12         | 616   | 40.6        | 283    | 1  | ISPE_ECOLI  |
| 13         | 616   | 40.6        | 295    | 1  | ISPE_VIBCH  |
| 14         | 615   | 40.6        | 281    | 1  | ISPE_NEIME  |
| 15         | 606.5 | 40.0        | 313    | 1  | ISPE_NAEIN  |
| 16         | 599.5 | 39.5        | 295    | 1  | ISPE_XANAC  |
| 17         | 584   | 38.5        | 295    | 1  | ISPE_XANCP  |
| 18         | 580.5 | 38.3        | 298    | 1  | ISPE_XYLEA  |
| 19         | 513.5 | 33.9        | 304    | 1  | ISPE_WIGER  |
| 20         | 493.5 | 32.6        | 286    | 1  | ISPE_WIGER  |
| 21         | 484.5 | 32.0        | 292    | 1  | ISPE_BUCAL  |
| 22         | 354   | 23.4        | 296    | 1  | ISPE_BUCAP  |
| 23         | 332   | 21.9        | 283    | 1  | ISPE_CHUTE  |
| 24         | 332   | 21.9        | 283    | 1  | ISPE_STRA3  |
| 25         | 327.5 | 21.6        | 293    | 1  | ISPE_STRA5  |
| 26         | 324   | 21.4        | 289    | 1  | ISPE_BACSU  |
| 27         | 323   | 21.3        | 291    | 1  | ISPE_LISIN  |
| 28         | 322.5 | 21.3        | 299    | 1  | ISPE_BRUME  |
| 29         | 321   | 21.2        | 282    | 1  | ISPE_STRUM  |
| 30         | 320   | 21.1        | 293    | 1  | ISPE_LISMO  |
| 31         | 317   | 20.9        | 287    | 1  | ISPE_BACHD  |
| 32         | 314   | 20.7        | 288    | 1  | ISPE_CLOPE  |
| 33         | 313.5 | 20.7        | 299    | 1  | ISPE_BRUSU  |

## ALIGNMENTS

| RESULT 1 | ISPE_PSEAE   | STANDARD | PRT | 282 AA |
|----------|--|----------|-----|--------|
| AC       | ISPE_PSEAE   |          |     |        |
| AC       | P42805   |          |     |        |
| DT       | 01-NOV-1995 (Rel. 32, Created)   |          |     |        |
| DT       | 16-OCT-2001 (Rel. 40, Last sequence update)  |          |     |        |
| DT       | 28-FEB-2003 (Rel. 41, Last annotation update)  |          |     |        |
| DE       | 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)   |          |     |        |
| DE       | 4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase)  |          |     |        |
| GN       | ISPE OR P44669   |          |     |        |
| OC       | Pseudomonas aeruginosa.  |          |     |        |
| OC       | Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  |          |     |        |
| OC       | Pseudomonadaceae; Pseudomonas.   |          |     |        |
| OX       | NCBI_TaxID=287;  |          |     |        |
| RN       | [1]  |          |     |        |
| RP       | SEQUENCE FROM N.A.   |          |     |        |
| RC       | STRAIN=ATCC 15692 / PAOI;  |          |     |        |
| RC       | MEDLINE=20437337; PubMed=10984043;   |          |     |        |
| RA       | Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  |          |     |        |
| RA       | Hickey M.J., Brinkman F.S.L., Hutmagle W.O., Kowalik D.V., Lagrou M.,  |          |     |        |
| RA       | Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  |          |     |        |
| RA       | Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  |          |     |        |
| RA       | Smith K.A., Spencer D.H., Wong G.K.-S., Mu Z., Paulsen I.T.,   |          |     |        |
| RA       | Reiser J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;   |          |     |        |
| RT       | "Complete genome sequence of Pseudomonas aeruginosa PAOI, an   |          |     |        |
| RT       | opportunistic pathogen."   |          |     |        |
| RT       | Nature 406:959-964(2000).  |          |     |        |
| [2]      |  |          |     |        |
| RC       | SEQUENCE OF 1-221 FROM N.A.  |          |     |        |
| RC       | STRAIN=ATCC 15692 / PAOI;  |          |     |        |
| RX       | MEDLINE=95189718; PubMed=7883699;  |          |     |        |
| RA       | Hungerer C., Troup B., Roemling U., Jahn D.;   |          |     |        |
| RT       | "Regulation of the hema gene during 5-aminolevulinic acid formation  |          |     |        |
| RT       | in Pseudomonas aeruginosa."  |          |     |        |
| RL       | Bacteriol. 177:1435-1443(1995).  |          |     |        |
| CC       | -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY  |          |     |        |
| CC       | GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY   |          |     |        |
| CC       | STIMULATORY)   |          |     |        |
| CC       | -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  |          |     |        |
| CC       | -erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-  |          |     |        |
| CC       | D-erythritol.  |          |     |        |
| CC       | -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth  |          |     |        |
| CC       | step.  |          |     |        |
| CC       | -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.  |          |     |        |
| CC       | This SWISS-PROT entry is copyright. It is produced through a collaboration   |          |     |        |
| CC       | between the Swiss Institute of Bioinformatics and the EMBL outstation in   |          |     |        |
| CC       | the European Bioinformatics Institute. There are no restrictions on its  |          |     |        |
| CC       | use by non-profit institutions as long as its content is in no way   |          |     |        |
| CC       | modified and this statement is not removed. Usage by and for commercial  |          |     |        |
| CC       | entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> |          |     |        |
| CC       | or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).  |          |     |        |
| CC       | -----  |          |     |        |
| DR       | EMBL, AE004880; AAC08056.1; --   |          |     |        |
| DR       | EMBL, X82071; CAAS7570.1; --   |          |     |        |
| DR       | FIR, F83063; F83063.   |          |     |        |

DR PIR: S49374; S49374.  
 DR HAMAP\_MF00061; -, 1.  
 DR InterPro: IPR006204; GMP kinase.  
 DR InterPro: IPR004424; ISPE.  
 DR Pfam: PF00288; GMP kinases; 1.  
 DR TIGRPFAM: TIGR00154; ISPE; 1.  
 DR Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;  
 KM Complete proteome.  
 FT NP\_BIND 95 105 ATP (POTENTIAL).  
 SQ SEQUENCE 282 AA; 30843 MW; 1418EBD4CF9438D3 CRC64;  
 Query Match 44.1%; Score 668; DB 1; Length 282;  
 Best Local Similarity 51.8%; Pred. No. 3.8e-53;  
 Matches 141; Conservative 36; Mismatches 95; Indels 0; Gaps 0;  
 QY 12 PAPAALNTMRTGRPRDGYHLQTVFQMLDCMLTFHPVDGKVTLRNPIGVPEDD 71  
 DB 8 PAPAALNTMRTGRPRDGYHLQTVFQMLDCMLTFHPVDGKVTLRNPIGVPEDD 67  
 QY 72 LTVRAANLKSHTGCVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 131  
 DB 68 LTVRAANLKSHTGCVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 127  
 QY 132 MDGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 191  
 DB 128 AALGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 187  
 QY 192 RNSAVVTMSDF-----LAGDRNDCSEVCTLRPVKADALALCYAEALITGACVPAFCNK 251  
 DB 188 RNSAVVTMSDF-----LAGDRNDCSEVCTLRPVKADALALCYAEALITGACVPAFCNK 247  
 QY 252 EDAESALBEGDKRMVLVFLAKGLNQSALYKLE 283  
 DB 248 EDAESALBEGDKRMVLVFLAKGLNQSALYKLE 279  
 RESULT 2  
 ISPE\_RALSO STANDARD; PRT; 289 AA.  
 AC Q8Y2E0;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)  
 GN (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).  
 OS ISPE OR IPK OR RSC0396 OR RSC03325.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OC NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangent S.,  
 RA Arlat M., Billaut A., Brotier P., Camus J.C., Catalicio L.,  
 RA Chandler M., Choise N., Claudi-Kernard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schex T.,  
 RA Sauter P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RT Nature 415:497-502(2002).  
 CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy  
 group of 4-diphosphocytidylyl-2-C-methyl-D-erythritol (By  
 similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  
 erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
 methyl-D-erythritol.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway, fourth  
 step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 DR ZMEP: AL646052; CAD13924.1; -.  
 DR HAMAP\_MF00061; -, 1.  
 DR InterPro: IPR006204; GMP kinase.  
 DR InterPro: IPR004424; ISPE.  
 DR Pfam: PF00288; GMP kinases; 1.  
 DR TIGRPFAM: TIGR00154; ISPE; 1.  
 DR Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;  
 KM Complete proteome.  
 FT NP\_BIND 99 109 ATP (POTENTIAL).  
 SQ SEQUENCE 289 AA; 31257 MW; A046E75006587A1 CRC64;  
 Query Match 43.9%; Score 665; DB 1; Length 289;  
 Best Local Similarity 51.3%; Pred. No. 7.3e-53;  
 Matches 140; Conservative 34; Mismatches 93; Indels 6; Gaps 2;  
 QY 12 PAPAALNTMRTGRPRDGYHLQTVFQMLDCMLTFHPVDGKVTLRNPIGVPEDD 71  
 DB 12 PAPAALNTMRTGRPRDGYHLQTVFQMLDCMLTFHPVDGKVTLRNPIGVPEDD 71  
 QY 72 LTVRAANLKSHTGCVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 131  
 DB 72 LTVRAANLKSHTGCVRGVCIDIEKNLPMGGGIGGSSDAATTLVNLRLMGLSKREL 131  
 QY 132 MDGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 191  
 DB 132 MDGLRLGADVPVFGCSAMGEVSEDLQAITLPBQMFVIIKPDCHVNTGEIIFSAENLT 191  
 QY 192 RNSAVVTMSDF-----LAGDRNDCSEVCTLRPVKADALALCYAEALITGACVPAFCNK 245  
 DB 192 RNSAVVTMSDF-----LAGDRNDCSEVCTLRPVKADALALCYAEALITGACVPAFCNK 251  
 QY 246 AQCCKEPAESALBEGDKRMVLVFLAKGLNQSAL 278  
 DB 252 AQCCKEPAESALBEGDKRMVLVFLAKGLNQSAL 284  
 RESULT 3  
 ISPE\_YERPE STANDARD; PRT; 299 AA.  
 AC Q8ZEX1;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)  
 GN (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).  
 OS ISPE OR IPK OR YPO2014 OR Y2293.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OC NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Churcher C., Mungall K.L.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Baeham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltham T., Hamlin N., Holtroyd S., Jagers K., Kariyasev A.V.,  
 RA Leather S., Moulton S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague";  
 RT Nature 413:523-527(2001).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=K1M5 / Biovar Mediaevalis;

RA MEDLINE=22137663; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G., III, Boutin A., Mayhew G.F., Liss P.,  
RA Perna N.T., Rose D.J., Mau E., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Linder J.E., Brubaker R.R., Plano G.V.,  
RA Straley S.C., McDonough K.A., Niles M.L., Natson J.S., Blattner F.R.,  
RA Perry R.D.,  
RA "Genome sequence of Yersinia pestis KIM";  
RA J. Bacteriol. 184:4601-4611(2002).  
CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy  
CC group of 4-diphosphocytidyl-2C-methyl-D-erythritol (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
CC methyl-D-erythritol.  
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth  
CC step.  
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AJ414151; CAC90828.1; -;  
CC EMBL; AE013832; AAM85852.1; -;  
CC PIR; AH0245; AH0245.1;  
CC HAMAP; MF\_00061; 1.  
CC Interpro; IPR006204; GMP\_Kinase.  
CC Interpro; IPR004424; Ispe.  
CC Pfam; PF00288; GMP\_Kinases; 1.  
CC TIGRfams; TIGR00154; ispe; 1.  
CC Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;  
CC Complete proteome.  
CC NP\_BIND 106 116 ATP (POTENTIAL).  
CC SEQUENCE 299 AA; 32675 MW; E86651DA68E0D658 CRC64;  
SQ  
Query Match 43.0%; Score 652.5; DB 1; Length 299;  
Best Local Similarity 49.8%; Pred. No. 1e-51;  
Matches 137; Conservative 42; Mismatches 91; Indels 5; Gaps 3;  
QY 10 RNPAPKALMLRITRRPDGTHLQTVFOMDLCDMLTFHVDDGRVTLRNPISGVEQ 69  
DB 14 KMPSPAKMLFYITGQRADYHQLTFLQFDYQGLTIFBRDQRLPLPIAGVNE 73  
QY 70 DDLTVRAANLKSHTG---CVRGVCIDIERKLPWGGLGSSSDAATTVANLMLGSL 126  
DB 74 QNLIVRAAKMLQKHPENTVPRGADISIKCLPMGGGLGSSSNATVVALNLMGGL 133  
QY 127 SKRELMDLGLRLGADVVFVFGSGAMGSEVSEDLQATLPEQWVLIKPDCHVTGEIFS 186  
DB 134 TBEQLADLGLTIGADYVFRGHAAPAEIGEKLPFAEVEKRWYIVIHGVNIPIPIIFS 193  
QY 187 AENLTPNSAVVMSDPLAGDNINCKSEVCKLRPYKADIDALLCYAARLITGACVFA 246  
DB 194 DPELKNYTRIPALILSTPYANICEPIARKRPREVEQSLWLEAPBRLITGACVFA 253  
QY 247 QFCNKEDASALEGLKDRML-VFLAKGLNQSALYK 280  
DB 254 EFDTESSARQVL-STAPEWLHGFAVGAVVSPHLR 287  
RESULT 4  
ID ISPE SHEON STANDARD; PRT; 284 AA.  
AC Q8BAR0;  
DT 15-SEP-2003 (Rel. 42; Created)  
DT 15-SEP-2003 (Rel. 42; Last sequence update)  
DT 15-SEP-2003 (Rel. 42; Last annotation update)  
DE 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)  
DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).

GN ISPE OR S03836.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=22297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Neilson K.E., Gaidos E.J., Neilson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
RA Keyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Impram M., Lee K., Berry K., Lee C.,  
RA Mueller J., Khouli H., Gill J., Utterback T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.E., Fraser C.M.,  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
RT Shewanella oneidensis".  
RL Nat. Biotechnol. 20:1118-1123(2002).  
CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy  
CC group of 4-diphosphocytidyl-2C-methyl-D-erythritol (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-  
CC methyl-D-erythritol.  
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth  
CC step.  
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE015816; AAN56813.1; -;  
CC TIGR; S03836; -;  
CC HAMAP; MF\_00061; -; 1.  
CC Pfam; PF00288; GMP\_Kinases; 1.  
CC TIGRfams; TIGR00154; ispe; 1.  
CC Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;  
CC Complete proteome.  
CC NP\_BIND 98 108 ATP (POTENTIAL).  
CC SEQUENCE 284 AA; 30721 MW; F5A22B814242496 CRC64;  
SQ  
Query Match 42.5%; Score 644.5; DB 1; Length 284;  
Best Local Similarity 49.1%; Pred. No. 5.1e-51;  
Matches 135; Conservative 41; Mismatches 98; Indels 1; Gaps 1;  
QY 11 WPAKAKMLKRLIRGRRPDGTHLQTVFOMDLCDMLTFHVDDGRVTLRNPISG-VPEQ 69  
DB 9 WPAKAKMLFLHNGRRSDGTHLQTLFQFVDCDQDFRTDPELILHSTMAVAVDS 68  
QY 70 DDLTVRAANLKSHTGCVRGVCIDIERKLPWGGLGSSSDAATTVANLMLGSLKSR 129  
DB 69 DNLIRPAKSLQOATIGRNGAIEWLDRLEPWGGIGSSSDAATTVANLMLMTQUSHD 128  
QY 130 ELMDLGLRLGADVVFVFGSGAMGSEVSEDLQATLPEQWVLIKPDCHVTGEIFS 189  
DB 129 ELAIGLGLAGDILVFIHGFAAPAGVGERLQAVNPRLMVLVIAPDHVGTAIVFQDPL 188  
QY 190 LTRNSAVVMSDPLAGDNINCKSEVCKLRPYKADIDALLCYAARLITGACVFA 249  
DB 189 LPRNTPLKGLDPTLISQPMANDQGLVYSKPPQVAKAGWLEAPBRLITGACVFA 248  
QY 250 NKEDASALEGLKDRMLVFLAKGLNQSALYKLEQ 284  
DB 249 SQOALALALALPSDMGQFAKGNISPLIYRLNR 283

| ID                        | ISPE_VIBVU  | STANDARD:                         | PRT: | 290 AA. |
|---------------------------|---|-----------------------------------|------|---------|
| AC                        | Q8PFF6  |                                   |      |         |
| DT                        | 15-SEP-2003   | (Rel. 42, Created)                |      |         |
| DT                        | 15-SEP-2003   | (Rel. 42, Last sequence update)   |      |         |
| DT                        | 15-SEP-2003   | (Rel. 42, Last annotation update) |      |         |
| DE                        | 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)                                    |                                   |      |         |
| DE                        | 4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).  |                                   |      |         |
| ISPE                      | OR VY10256.   |                                   |      |         |
| OS                        | Vibrio vulnificus.  |                                   |      |         |
| OC                        | Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;   |                                   |      |         |
| OC                        | Vibrionaceae; Vibrrio.  |                                   |      |         |
| OX                        | NCBI_TaxID=672;   |                                   |      |         |
| RA                        | SEQUENCE FROM N.A.  |                                   |      |         |
| RA                        | STRAIN=CMCP6;   |                                   |      |         |
| RA                        | Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,   |                                   |      |         |
| RA                        | Choy H.E.;  |                                   |      |         |
| RT                        | "Complete genome sequence of Vibrio vulnificus CMCP6.";   |                                   |      |         |
| RU                        | Submitted (DDBJ/EMBL/GenBank/DBJ databases).  |                                   |      |         |
| CC                        | - FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy                                       |                                   |      |         |
| CC                        | group of 4-diphosphocytidyl-2C-methyl-D-erythritol (by  |                                   |      |         |
| CC                        | similarity).  |                                   |      |         |
| CC                        | - CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-                                       |                                   |      |         |
| CC                        | erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-   |                                   |      |         |
| CC                        | methyl-D-erythritol.  |                                   |      |         |
| CC                        | - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth   |                                   |      |         |
| CC                        | step.   |                                   |      |         |
| CC                        | - SIMILARITY: BELONGS TO THE ISPE FAMILY.   |                                   |      |         |
| CC                        | -----   |                                   |      |         |
| CC                        | This SWISS-PROT entry is copyright. It is produced through a collaboration                                |                                   |      |         |
| CC                        | between the Swiss Institute of Bioinformatics and the EMBL outstation -                                   |                                   |      |         |
| CC                        | the European Bioinformatics Institute. There are no restrictions on its                                   |                                   |      |         |
| CC                        | use by non-profit institutions as long as its content is in no way com-                                   |                                   |      |         |
| CC                        | modified and this statement is not removed. Usage by and for commercial                                   |                                   |      |         |
| CC                        | entities requires a license agreement (see <a href="http://www.isb-sib.ch/">http://www.isb-sib.ch/</a> ). |                                   |      |         |
| CC                        | or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).                         |                                   |      |         |
| CC                        | -----   |                                   |      |         |
| DR                        | EMBL; AE016797; AAC08792.1; -   |                                   |      |         |
| DR                        | HAMAP; MF 00061; -; 1.  |                                   |      |         |
| DR                        | Pfam; PF00288; GMP kinases; 1.  |                                   |      |         |
| DR                        | TIGRfam; TIGR00154; ispe; 1.  |                                   |      |         |
| KM                        | Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;   |                                   |      |         |
| KM                        | Complete proteome.  |                                   |      |         |
| FT                        | NP_BIND 96 106  |                                   |      |         |
| FT                        | SEQUENCE 290 AA: 31657 MW; 27161EC31E1EDB8D CRC64;  |                                   |      |         |
| SO                        |   |                                   |      |         |
| Query Match               | 42.3%; Score 642; DB 1; Length 290;   |                                   |      |         |
| Best Local Similarity     | 48.9%; Pred. No. 6.9e-51;   |                                   |      |         |
| Matches 133; Conservative | 43; Mismatches 96; Indels 0; Gaps 0;  |                                   |      |         |
| DY                        | 11 WPAPKNTLMRTIRRPDGYHLLQTVQMDLCMLTFHPYDGRVTLKRPISGVEPD 70  |                                   |      |         |
| DB                        | 8 WPSPKNLWFIYINRTDNGYHELTLPQPLDHGQLITRANDSGHITLTPDIYDLPEQ 67  |                                   |      |         |
| DY                        | 71 DLTVAAANILKSHTCVARGVCIDIRKNI.PMGGLGGSSDAATTLYVINRLWGLSKRE 130  |                                   |      |         |
| DB                        | 68 NLIMWAAVALQKRGCTGLANIHMLKLI.PMGGLGGSSNAATLVALNPLMQGLSDDE 127   |                                   |      |         |
| DY                        | 131 LMDGLRLGADVPRVYFEGCSAMGEGVEDLQATILPEQMFYILKDDCVNNGEIPSAENL 190  |                                   |      |         |
| DB                        | 128 LADIGLKLGAADVPRVYFEGHAAFAEYVGRGLTPAQBEEKWYLVPRDVAIATVDIPTHQL 187                                      |                                   |      |         |
| DY                        | 191 TRNSAVVTMSDFLGDNRNDSSEVVCXYRPAKDAILCYAEARLETGTGACVPAQFCN 250  |                                   |      |         |
| DB                        | 188 TRNIPKRSLETLLDSEYGNDEKILRMHPKVDKQLSMILQYAPBRLTGTSQVFAERN 247  |                                   |      |         |
| DY                        | 251 KDAESALEGLKDRWLVLAKNLSALYKYL 282  |                                   |      |         |
| DB                        | 248 RSEAESILAQLSDNVSAPVAQGRNLSPLKETL 279  |                                   |      |         |

```

RESULT 6
ID ISPE_VIBRA STANDARD; PRT; 295 AA.
AC Q87RNT;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)
DB (-4-)diphospho-5'-diphospho)-2-C-methyl-D-erythritol kinase).
GN ISPE OR VP0740.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 221063 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=1262073;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nejlina M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yaeguchi T., Honda T., Shinagawa H., Hattori M., Ikeda T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC -I- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
CC group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (By
CC similarity).
CC -I- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol.
CC -I- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -I- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF005075; BACS9003.1; -.
DR HAMAP; MF 00061; -. 1.
KW Transferrase; Kinase; Isoprene biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 101 111 ATP (POTENTIAL).
SQ SEQUENCE 295 AA; 32497 MW; EF77FF8973BDC7ED CRC64;
Query Match 42.0%; Score 636; DB 1; Length 295;
Best Local Similarity 48.0%; Pred. No;
Matches 132; Conservative 46; Mismatches 97; Indels 0; Gaps 0;
QY 10 RWPARAKLNLMTKRTGRBPQSTHLLOVFWQLDCMLTFHPVDGVTLENPLSGVEQ 69
DB 12 RWPSPARKINTFLYNINERTENGTHLETLFQFVDHGDGLFIQANHSQDVITSPETEGVFLLQ 71
QY 70 DDLTVBARANKLSHTQCVARGVIDIEBKRLPKGGIGAGSSSDAATLVYLNLWLGISKR 129
DB 72 DNLIWKAAITALQNYAHKSPFGNHIELHKVLLPWGGIGGGSSNAATLVVALNTLMQNLITDD 131
QY 130 ELMDLGLRLGADVVPFPFGCSAMBEGISSEDLQALTEPQWTVIIKKPDCHVNTGSIFSEN 189
DB 132 ELAIKIGKLGLADGVPPFRGFAPAEVGGEKSLPANPEEKMTLVLRPNVSITAFIRHPD 191
QY 190 LTRNSAVVMSDFLAGNRNDSCSEVCKLYRPVDAIDALCYEARLTGTGACVFAQC 249
DB 192 LTRNTTPRKDELTTLNARSVNDCEKIIVMAYEVDKQSLTWLQYAPSRLTGTGSCVPAEFS 251
QY 250 NKEDAASLBGIKDKRWLVFLAKGINQSALTGLEQ 284
DB 252 SKSEAETTLAQSLDKVASFAVAQGNIIPLPKEKTAA 286

```





CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; W77236; AAA27195.1; -  
 CC EMBL; AE008779; AAL20694.1; -  
 CC EMBL; M19488; -; NOT\_ANNOTATED\_CDS.  
 CC PIR; S27732; S27732.  
 CC StyGene; SGI00417; ispe.  
 CC HAMAP; MF\_00061; -; 1.  
 CC InterPro; IPR006204; GHMP\_kinase.  
 CC InterPro; IPR004424; ispe.  
 CC Pfam; PF00288; GHMP\_kinases; 1.  
 CC TIGRFAMs; TIGR00154; ispe; 1.  
 CC Transferase; Kinase; Isoprene biosynthesis; ATP-binding;  
 CC Complete proteome.  
 CC NP\_BIND 99 109 ATP (POTENTIAL).  
 CC FT CONFLICT 252 252 S -> L (IN REF. 3).  
 CC FT SEQUENCE 283 AA; 30883 MW; E33EAF956D12A5BD CRC64;  
 CC -----  
 CC Query Match 41.5%; Score 629; DB 1; Length 283;  
 CC Best Local Similarity 50.2%; Pred. No. 1.3e-49;  
 CC Matches 140; Conservative 35; Mismatches 96; Indels 8; Gaps 4;

QY 11 WPAPAKLIMLRITGRPDGYHLLQTVFQMLDCWLTGHPVDDGAVTLRNISGVPEOD 70  
 DB 5 WPSPAKLNFELYTGQADGVHTLQTLFQFLDYGDTLHLEPHDDGHIILTFVNGVEMED 64  
 QY 71 DLTVRANLILKSHITGCV---RGVICIEKTLPMGGGLGGSSAAATTVLARIKGL 124  
 DB 65 NLIVRARLIMKRVASSGRPLPGSGADISTEKRLPMGGGLGGSSAAATTVLALNHLWQC 124  
 QY 125 GSKRELMDLGLRLGADVFFVFGCSAMGEVSEBDQAITLPEQWFIILKPDCHVNTGEI 184  
 DB 125 GLSIDELATLGLTGLADVFFVFGGHAFAEGVGEILITPVNPEKVLVHNPVSIPTPYI 184  
 QY 185 FSEKENTRISAVVTMSDFLAGDNKNDCEVCKLTPVYDAIDALLCYAEALITGACV 244  
 DB 185 FPDQPLPNTPKRSITDLTKCEFSNDEVIARRFRFVDAALSMLELVAPSRLTGACV 244  
 QY 245 PAQFCNKEDAESALGLKDRWL-VFLAKGLNOSALYKYL 282  
 DB 245 FAFOTESCARGVLEGAPE-WLNAFAKGVNLSPLHREL 282

RESULT 9  
 ISPE NEIMA STANDARD; PRT; 281 AA.  
 ID ISPE NEIMA  
 AC 09UK8;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)  
 DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase)  
 GN ISPE OR NMA1092.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 NCBI\_TaxID=56399;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=22491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies K.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,  
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

RA Whitehead S., Spratt B.G., Barrell B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 RT meningitidis 22491.";  
 RL Nature 404:502-506(2000).  
 CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY  
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY  
 CC SIMILARITY)  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-  
 CC D-erythritol.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth  
 CC step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AL162755; CAB84355.1; -  
 CC PIR; A61875; A61875.  
 CC HAMAP; MF\_00061; -; 1.  
 CC InterPro; IPR006204; GHMP\_kinase.  
 CC InterPro; IPR004424; ispe.  
 CC Pfam; PF00288; GHMP\_kinases; 1.  
 CC TIGRFAMs; TIGR00154; ispe; 1.  
 CC Transferase; Kinase; Isoprene biosynthesis; ATP-binding;  
 CC Complete proteome.  
 CC NP\_BIND 98 108 ATP (POTENTIAL).  
 CC FT SEQUENCE 281 AA; 31273 MW; E8A9AB81CCB40B80 CRC64;  
 CC -----  
 CC Query Match 41.2%; Score 624.5; DB 1; Length 281;  
 CC Best Local Similarity 46.7%; Pred. No. 3.3e-49;  
 CC Matches 128; Conservative 51; Mismatches 88; Indels 7; Gaps 2;

QY 9 ERWPAPAKLIMLRITGRPDGYHLLQTVFQMLDCWLTGHPVDDGAVTLRNISGVPE 68  
 DB 8 QAPAPAKLIMLRITGRPDGYHLLQTVFQMLDCWLTGHPVDDGAVTLRNISGVPE 67  
 QY 69 QDDITVRANLILKSHITGCV---RGVICIEKTLPMGGGLGGSSAAATTVLARIKGL 126  
 DB 68 EADLSYRPAASLLQYARTPTVATLWLDKLIPTGAGLGGSSDAATVLLVLRMWCGLTQ 127  
 QY 129 RELMDLGLRLGADVFFVFGCSAMGEVSEBDQAITLPEQWFIILKPDCHVNTGEI 188  
 DB 128 RQIDSGALGLADVFFVFGGHAFAEGVGEILITPVNPEKVLVHNPVSIPTPYI 187  
 QY 189 NITRNSAVVTMSDFLAGDN---ENDCEVYCKLTPVYDAIDALLCYAEALITGACV 244  
 DB 188 GTRNSASISIMPTF---QNLQPFRRDMQAVFKEVPEWKAYSLSRGFALMTGSGACV 244  
 QY 245 PAQFCNKEDAESALGLKDRWL-VFLAKGLNOSAL 278  
 DB 245 FTACODRNSAINTYKQVSDIVYATVLAISLHREL 278

RESULT 10  
 ISPE PASMU STANDARD; PRT; 295 AA.  
 ID ISPE PASMU  
 AC P57833;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)  
 DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase)  
 GN ISPE OR PM0245.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 NCBI\_TaxID=747;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.,
RT "Complete genomic sequence of Pasteurella multocida pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
CC D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR HAMAP; A8006059; AAK02329.1; -.
CC DR InterPro; IPR006204; GMP_Kinase.
CC DR InterPro; IPR004424; ISPE.
CC DR Pfam; PF00288; GMP_kinases; 1.
CC DR TIGRFAMs; TIGR00154; ISPE; 1.
CC DR Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
CC KM Complete proteome.
CC FT NP BIND 108 118 ATP (POTENTIAL).
CC SQ SEQUENCE 295 AA; 33064 MW; 5FE9FEB7C9ADB55 CRC64;

Query Match 41.2%; Score 624; DB 1; Length 295;
Best Local Similarity 45.4%; Pred. No. 3.9e-49;
Matches 124; Conservative 49; Mismatches 100; Indels 0; Gaps 0;

QY 10 RMPAPAKLNLMTLRTRGRRPGYHLQTVFQMLDCLDWLTHPHVDGKVTLRNPISGVPQ 69
DB 19 RFPAPKLNPLFYINGRQGGYHELOTLFQVFGDWLDEVEDNEICLTPELPSIKNE 78
QY 70 DDLTVRAANLMLKSHTGCVRGVCIIEKNLPMGGGLGGSSDAATTVLNLKALSKR 129
DB 79 DNLVYRAKXLIQKQTNALGANLTLKILPMGSELGGSSNATRALVALNLTQULSTR 138
QY 130 ELMDLGLRLGADVVPYFGCSAMGEGVSEDLQATLTPQWFIIVIKDCHVNTGEIISAEN 189
DB 139 QLAQGLMLGADVPIFVGHAAFAEYGEKITVCEPEREKYVVLKKNVSIATVFSDDP 198
QY 190 LTRSAVAVTMSDFLAGNNDGSEVCKLYRPVDAIDALCYABARHGTGAQVAPQC 249
DB 199 LINTPQOSLEQLLNQRYANDCEKVNINHPVEEIIHRLQVAPSRSLTGTGACVARERN 258
QY 250 NKEDASALEGLKDRMLVFLAKGINSALYKCL 282
DB 259 DEBSAQALAFOTIPKNYFGVAAQGLNKSPLANML 291

RESULT 11
ISPE_BCOL6 STANDARD; PRT; 283 AA.
AC Q8PI04;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CKM)
DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
GN ISPE OR C166.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

```

OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HL / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liu S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.U., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli O6."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -1- FUNCTION: Catalyzes the phosphorylation of the position 2 hydroxy
CC group of 4-diphosphocytidylyl-2C-methyl-D-erythritol (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-
CC methyl-D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR HAMAP; A8016759; AAN80131.1; -.
CC DR HAMAP; MF 00061; -.
CC DR Pfam; PF00288; GMP_kinases; 1.
CC DR TIGRFAMs; TIGR00154; ISPE; 1.
CC DR Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
CC KM Complete proteome.
CC FT NP BIND 99 109 ATP (POTENTIAL).
CC SQ SEQUENCE 283 AA; 30973 MW; E0327295B5EDDFC CRC64;

Query Match 40.9%; Score 620; DB 1; Length 283;
Best Local Similarity 47.9%; Pred. No. 8.4e-49;
Matches 135; Conservative 40; Mismatches 95; Indels 12; Gaps 4;

QY 10 RMPAPAKLNLMTLRTRGRRPGYHLQTVFQMLDCLDWLTHPHVDGKVTLRNPISGVPQ 69
DB 4 QMPSPKLNPLFYITGCRADSGHTLQTLFQFLDYGDGTISLRDSDNILLTPVEGVEHE 63
QY 70 DDLTVRAANLMLKSHTGCVRGVCIIEKNLPMGGGLGGSSDAATTVLNLKALSKR 121
DB 64 DNLVYRAKXLIQKQTNALGANLTLKILPMGSELGGSSNATRALVALNLTQULSTR 121
QY 122 KGLGSKREIMDLGLRLGADVVPYFGCSAMGEGVSEDLQATLTPQWFIIVIKDCHVNT 181
DB 122 WOCGLSMDLAMEGLTGLADVVPYFGHAAFAEYGEKITVCEPEREKYVVLKKNVSIATVFSDDP 181
QY 182 GEIPISEAVTMSDFLAGNNDGSEVCKLYRPVDAIDALCYABARHGTGAQVAPQC 241
DB 182 PVTFQDPELPRNTPRISITETLLKCFPSNDCEVIARKRRREDAVAISWLEVAFSLRTGIG 241
QY 242 ACVFAEFTESEARQVLBOAPE-WLNGEVAKGVNLSPIRRAM 282
DB 242 ACVFAEFTESEARQVLBOAPE-WLNGEVAKGVNLSPIRRAM 282

RESULT 12
ISPE_BCOL1 STANDARD; PRT; 283 AA.
AC F24209;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CKM)

```

DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).  
 GN ISPE OR IPK OR B1208 OR Z1979 OR EGS1713.  
 OS *Escherichia coli*, and  
 OS *Escherichia coli* 0157:H7.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; *Escherichia*.  
 RX NCBI\_TaxID=562, 83334.  
 RN [1]  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=20040616; PubMed=10570138;  
 RA Lange B.M., Crocrau R.;  
 RT "Isopentenyl diphosphate biosynthesis via a mevalonate-independent  
 RT pathway: isopentenyl monophosphate kinase catalyzes the terminal  
 RT enzymatic step.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13714-13719(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RC STRAIN=K12 / W3110;  
 RA Kuzuyama T., Takagi M., Kaneda K., Watanabe H., Dairi T., Seto H.;  
 RT "Studies on the nonmevalonate pathway: Conversion of 4-(cytidine  
 RT 5'-diphospho)-2-C-methyl-D-erythritol to its 2-phospho derivative by  
 RT 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase.";  
 RL Tetrahedron Lett. 41:2925-2928(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RX MEDLINE=20122571; PubMed=10655484;  
 RA Luettgen H., Rohdich F., Herz S., Wungsihaweekul J., Hecht S.,  
 RA Schuhr C.A., Fellemeier M., Sagner S., Zenk M.H., Bacher A.,  
 RA Benschel W.;  
 RT "Biosynthesis of terpenoids: Ycbh protein of *Escherichia coli*  
 RT phosphoxylates the 2-hydroxy group of 4-diphosphocytidylyl-2C-methyl-D-  
 RT erythritol.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1062-1067(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=93171869; PubMed=7679718;  
 RA Post D.A., Hove-Jensen B., Switzer R.L.;  
 RT "Characterization of the hemA-prs region of the *Escherichia coli* and  
 RT *Salmonella typhimurium* chromosomes: identification of two open  
 RT reading frames and implications for prs expression.";  
 RL J. Gen. Microbiol. 139:259-266(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93051347; PubMed=1427085;  
 RA Ikemi M., Murakami K., Hashimoto M., Murooka Y.;  
 RT "Cloning and characterization of genes involved in the biosynthesis  
 RT of delta-aminolevulinic acid in *Escherichia coli*.";  
 RL Gene 121:127-132(1992).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Remmler P., Woletschlaeger M., Strohmaier H.;  
 RT Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Coliello-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1233-1238(1997).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Ikemoto K., Inada H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,

RA Saito G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horuchi T.;  
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., DiStefano E.T., Potamistos K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* 0157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Teshi K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 RT 0157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 9:11-22(2001).  
 CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY  
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL. PHOSPHORYLATES  
 CC ISOPENTENYL PHOSPHATE AT LOW RATES. ALSO ACTS ON ISOPENTENYL, AND,  
 CC MUCH LESS EFFICIENTLY, DIMETHYLLALYL ALCOHOL. DIMETHYLLALYL,  
 CC MONOPHOSPHATE DOES NOT SERVE AS A SUBSTRATE.  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-  
 CC D-erythritol.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth  
 CC step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.  
 CC -1- CAUTION: Was originally (Ref.1) thought to be an isopentenyl  
 CC monophosphate kinase.  
 CC -1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N- AND  
 CC C-TERMINUS DUE TO FRAMESHIFTS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, AJ249325; CAB64963.1; -  
 CC EMBL, AF179284; AAF13867.1; -  
 CC EMBL, AF216300; AAF29530.1; -  
 CC EMBL, AB037116; BA344247.1; -  
 CC EMBL, W77237; AA24434.1; -  
 CC EMBL, D10264; BAA01106.1; ALT\_FRAME.  
 CC EMBL, U18555; AAC43434.1; -  
 CC EMBL, AE000219; AAC74292.1; -  
 CC EMBL, D90756; BAA36066.1; -  
 CC EMBL, AE005318; AAG56066.1; -  
 CC EMBL, AP002556; BAB35136.1; -  
 CC PIR, A90843; A90843.  
 CC PIR, B47706; B47706.  
 CC ECoGene: EG11294; 1spB.  
 CC HAMAP: MF\_00061; -; 1.  
 CC InterPro: IPR006204; GMP\_kinase.  
 CC InterPro: IPR004424; 1spB-  
 CC Pfam: PF00289; GMP\_kinases; 1.  
 CC TIGRFAMs: TIGR00154; 1spB; 1.  
 CC Transferase; Kinase; isoprene biosynthesis; ATP-binding;  
 CC Complete proteome.

FT NP BIND 99 109 ATP (POTENTIAL).  
 SQ SEQUENCE 283 AA; 30925 MW; 59A2921FA05D1B1 CRC64;  
 Query Match 40.6%; Score 616; DB 1; Length 283;  
 Best Local Similarity 47.9%; Pred. No. 1.9e-48;  
 Matches 135; Conservative 39; Mismatches 96; Indels 12; Gaps 4;  
 QY 10 RWPAPAKNLIMLRITGRPRPGYHLQTVFQMLDLCMLTFHPVDDGRVTLRNISGVPEQ 69  
 DB 4 QWSPKAKNLFLYITGRADGHTLTQLTQFDYDGTITSELRDDDDILRLPVEGVHE 63  
 QY 70 DDLTVAANL-----KSHTGCVRGVCIDIEKRLPMGGGLGGSSDAATLVVNLRL 121  
 DB 64 DNLIVRAARLIMKTRADSGRLPTG--SGANISIDKRLPMGGGLGGSSNAATLVVNLRL 121  
 QY 122 WGLGSLRELMDLGLGADVPVFGCSAMGSGVEDLQAITLPQCVIIRKDCVNT 181  
 DB 122 WCGGLSMDLELAEMGLTLGADVVPVFGHAFAGVGELITVPDPKRYLVAAHPSVIF 181  
 QY 182 GEIFSAENLIRNSAVVMTSDPLAGDNRNDCSEVVCCKLYRPVKAIDALLCYAEARLTGTG 241  
 DB 182 PVIFKDELPFRNTPKRSIEFTLKCEFSNDCEVIAKRFREVDVLSMLLEAPSLTGTG 241  
 QY 242 ACVPAQCKNEDKESALBGLKRWL-VFLAKGLNQSALYKRL 282  
 DB 242 ACVFAEDTSEARQVLEQAP-MLNGFVAKGALNPLHRAM 282  
 RESULT 13  
 ID ISPE\_VIBCH STANDARD; PRT; 295 AA.  
 AC Q9XQ23;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)  
 DS (4-cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).  
 GN ISPE OR VC2182.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 CC Vibrionaceae; Vibrio.  
 CX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson M.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Ueberlack T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RT Nature 406:477-483(2000).  
 RL NATURE 406:477-483(2000).  
 CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY  
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-  
 CC D-erythritol.  
 CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; fourth  
 CC step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: AE004289; AAF95327.1; -  
 DR PIR: B82109; B82109.  
 DR TIGR: VC2182; -  
 DR HAMAP: MF\_00061; -; 1.  
 DR InterPro: IPR006204; GMP Kinase.  
 DR InterPro: IPR004424; ISPE.  
 DR Pfam: PF00288; GMP Kinase; 1.  
 DR TIGRFAMs: TIGR00154; ispe; 1.  
 KM Transferrase, kinase; Isoprene biosynthesis; ATP-binding;  
 KM Complete proteome.  
 FT NP BIND 101 111  
 SQ SEQUENCE 295 AA; 31973 MW; DF82A14981710531 CRC64;  
 Query Match 40.6%; Score 616; DB 1; Length 295;  
 Best Local Similarity 47.1%; Pred. No. 2e-48;  
 Matches 128; Conservative 46; Mismatches 98; Indels 0; Gaps 0;  
 QY 11 WPAKAKNLIMLRITGRPRPGYHLQTVFQMLDLCMLTFHPVDDGRVTLRNISGVPEQ 70  
 DB 13 WSPKAKNLFLYITGRPRANGYHDLQTLFQFDHGDILTANSGNITLSPALADVALED 72  
 QY 71 DLTVAANLKSHTGCVAGVCLDIKRNPMGGGLGGSSDAATLVVNLRLMGLSKRE 130  
 DB 73 NLIVRAARLIMKTRADSGRLPTG--SGANISIDKRLPMGGGLGGSSNAATLVVNLRL 132  
 QY 131 LMDLGLRGADVPVFGCSAMGSGVEDLQAITLPQCVIIRKDCVNTGEIFSAENL 190  
 DB 133 LAEIGLALGADVVPVTRGPAFAEGVELSABERBNRYLVVRAVSLATDIFTHPL 192  
 QY 191 TNSAVVMTSDPLAGDNRNDCSEVVCCKLYRPVKAIDALLCYAEARLTGTGACVPAQCN 250  
 DB 193 MNTEPKRLASLITLPYRNDCEKIVRSIYPEDKQSLWLTQYAPSLTGTGSCVFAEFS 252  
 QY 251 KEDASALEGLKRWLVFLAKGLNQSALYKRL 282  
 DB 253 KRDQAVPRAQSLDNVLAFAVAGRNVSPLKRL 284  
 RESULT 14  
 ID ISPE\_NEIMB STANDARD; PRT; 281 AA.  
 AC Q9JZM4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)  
 DS (4-cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).  
 GN ISPE OR NMB0874.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 CC Neisseriaceae; Neisseria.  
 CX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / Serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson M.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,  
 RA Cotton M.D., Ueberlack T.R., Khouiri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 RT MC58.";  
 RL NATURE 406:1809-1815(2000).  
 CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY  
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2-C-METHYL-D-ERYTHRITOL (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-  
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-  
 CC D-erythritol.

```

CC D-erythritol.
CC -1- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A6002439; AAF41285.1;
DR PIR: B61149; B61149.
DR TIGR: M30874;
DR HAMAP: MF_00061; 1.
DR InterPro: IPR006204; GMP_kinase.
DR InterPro: IPR004424; ISPE.
DR Pfam: PF00288; GMP_kinases; 1.
DR TIGRFAMs: TIGR00154; ispe; 1.
DR Transferase: Kinase; Isoprene biosynthesis; ATP-binding;
KM Complete proteome.
FT NP_BIND 98 108 ATP (POTENTIAL).
SQ SEQUENCE 281 AA; 31315 MW; F8FA5216CC2B7703 CRC64;

Query Match 40.6%; Score 615; DB 1; Length 281;
Best Local Similarity 46.1%; Pred. No. 2,4e-48;
Matches 130; Conservative 53; Mismatches 91; Indels 8; Gaps 3;

QY 1 MDYAAGGKRWPAKNTLMRTTGRBPDGTHLLQTVFQMLDLCDWLTHPVDGAVTLR 60
DB 1 MNTADG-ROAFSAAPAKNTLMRTTGRBPDGTHLLQTVFQMLDLCDWLTHPVDGAVTLR 59
QY 61 NPISGVPEDDITVRANLTKSHTGCVRCVCIIEKNLPMGGGLGGSSDAATLVNLR 120
DB 60 NPVDGMPQEVDSLYRASLQKAKAPAGVEIWDKPIKIPAGLGGSSDAATLVNLR 119
QY 121 LMGIGLSKRELMGLGLGADVVPFVFGCSAMGVSSEDLQATLPEQWVILKPDCHYN 180
DB 120 WMCGGLTORGLDISGALGADVVPFIFGKNAFARIGDRIDEMDIPKQWTVIVKPVHVS 179
QY 181 TGIFFSEKENTRSAYVTMSDFLAGDN---RNDCEVYCKLYRPAKDAIDALCYAEAR 236
DB 180 TAIIFTHESLTRNSASIMPTF---QNLQPRNDQAVKPEVWKVAYSELSRYGVAL 236
QY 237 LMGTCACVFAQCNKEDAESALGLKDRMLVFLAKGINOSAL 278
DB 237 MTGSGACVFTACQDRNSAVTVYQVSLYEAVALAGLSKHP 278

RESULT 15
ISPE_HAEIN STANDARD; PRT; 313 AA.
ID ISPE_HAEIN STANDARD; PRT; 313 AA.
AC P4527;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CKM)
DE (4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
ISPE OR H11608.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

```

```

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sadek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
CC GROUP OF 4-DIPHOSPHOCYTIDYLYL-2C-METHYL-D-ERYTHRITOL (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
CC D-erythritol.
CC -1- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; fourth
CC step.
CC -1- SIMILARITY: BELONGS TO THE ISPE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32834; AAC3282.1;
DR TIGR: H11608;
DR HAMAP: MF_00061; 1.
DR InterPro: IPR006204; GMP_kinase.
DR InterPro: IPR004424; ISPE.
DR Pfam: PF00288; GMP_kinases; 1.
DR TIGRFAMs: TIGR00154; ispe; 1.
DR Transferase: Kinase; Isoprene biosynthesis; ATP-binding;
KM Complete proteome.
FT NP_BIND 113 123 ATP (POTENTIAL).
SQ SEQUENCE 313 AA; 34657 MW; 7A84BAKAL196821B CRC64;

Query Match 40.0%; Score 606.5; DB 1; Length 313;
Best Local Similarity 47.1%; Pred. No. 1.6e-47;
Matches 130; Conservative 45; Mismatches 100; Indels 1; Gaps 1;

QY 10 RMPAPAKNTLMRTTGRBPDGTHLLQTVFQMLDLCDWLTHPVDGAVTLNPSIGPE 68
DB 23 RPPSPAKNTLMRTTGRBPDGTHLLQTVFQMLDLCDWLTHPVDGAVTLNPSIGPE 68
QY 69 QDDLTVRAANTLKSHTGCVRCVCIIEKNLPMGGGLGGSSDAATLVNLRMGAGSK 128
DB 83 ENNLVYRAAKLQEBANITGLANITLHDKILPMGGVGGSSDAATLVNLRMGAGSK 142
QY 129 RRLMDGLRLGADVVPFVFGCSAMGVSSEDLQATLPEQWVILKPDCHYNVTSISAE 188
DB 143 DELARDGLTGLADVPIFVGHANAFAVGGEKITTCPEPAKRWVILKPDISTAVTFDDP 202
QY 189 NITRNSAVVTMSDFLAGDNNDCESEVCKLYRPAVDALDALCYAEARLTGTACVFAQF 248
DB 203 NLPRTTPKRSIAQLLSBPYKNDCEKVINHYSNEXALNTMLQVAPALITGTACVFAAF 262
QY 249 CNKEDAESALGLKDRMLVFLAKGINOSALYKKEEQ 284
DB 263 DHEAEQAVFRQKPEAFPGFVAKGINVSPHAKMKQ 298

Search completed: January 29, 2004, 15:50:40
Job time : 5.36681 secs

```

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 29, 2004, 15:44:14 : Search time 21.2436 Seconds  
(without alignments)  
3461.979 Million cell updates/sec

Title: US-09-941-947a-12  
Perfect score: 1516  
Sequence: 1 MEYAGWGERMPAPAKNLIM.....LVEFLAKGNQSLATYKLEEG 285

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: SP archaea:\*
- 2: SP bacteria:\*
- 3: SP fungi:\*
- 4: SP human:\*
- 5: SP invertebrate:\*
- 6: SP mammal:\*
- 7: SP\_mmc:\*
- 8: SP\_organelle:\*
- 9: SP\_phage:\*
- 10: SP\_plant:\*
- 11: SP\_rodent:\*
- 12: SP\_virus:\*
- 13: SP\_vertebrate:\*
- 14: SP\_unclassified:\*
- 15: SP\_virus:\*
- 16: SP\_bacteriap:\*
- 17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| 1          | 644.5 | 42.5        | 284    | 16    | Q8EARD Shewanella   |
| 2          | 642   | 42.3        | 290    | 16    | Q8DF66 Vibrio vuln  |
| 3          | 620   | 40.9        | 283    | 16    | Q8F104 Escherichia  |
| 4          | 513.5 | 33.9        | 304    | 16    | Q8D2K6 Wigglesworth |
| 5          | 332   | 21.9        | 283    | 16    | Q8E7K5 Streptococc  |
| 6          | 332   | 21.9        | 283    | 16    | Q8E245 Streptococc  |
| 7          | 321   | 21.2        | 282    | 16    | Q8D540 Streptococc  |
| 8          | 313.5 | 20.7        | 289    | 16    | Q8E2D0 Brucella su  |
| 9          | 302   | 19.9        | 282    | 16    | Q8CQ06 Staphylococ  |
| 10         | 280.5 | 18.5        | 311    | 16    | Q8DLJ1 Synechococc  |
| 11         | 280   | 18.5        | 280    | 16    | Q8E3J7 Oceanobactil |
| 12         | 274   | 18.1        | 401    | 10    | Q8E3G0 Oryza sativ  |
| 13         | 223.5 | 14.7        | 315    | 16    | Q8R0Z4 Corynebacte  |
| 14         | 204.5 | 13.5        | 316    | 16    | Q8E6I4 Bifidobacte  |
| 15         | 196   | 12.9        | 297    | 16    | Q8E2M8 Leptospira   |
| 16         | 178   | 11.7        | 141    | 16    | Q9J586 Chlamydia p  |

|    |       |     |     |    |                     |
|----|-------|-----|-----|----|---------------------|
| 17 | 134   | 8.8 | 265 | 16 | Q9PPN9 ureaplasma   |
| 18 | 125.5 | 8.3 | 312 | 16 | Q8E8T5 shewanella   |
| 19 | 124.5 | 8.2 | 301 | 17 | Q8PWJ3 methanosarc  |
| 20 | 119.5 | 7.9 | 286 | 17 | Q8TU21 methanopyru  |
| 21 | 116   | 7.7 | 813 | 4  | Q8N819 homo sapien  |
| 22 | 116   | 7.7 | 990 | 4  | Q96MT9 homo sapien  |
| 23 | 114   | 7.5 | 306 | 16 | Q8PLH7 xanthomonas  |
| 24 | 114   | 7.5 | 370 | 10 | Q8L7R2 arabidopsis  |
| 25 | 111.5 | 7.4 | 288 | 16 | Q8D2I2 streptococc  |
| 26 | 110   | 7.3 | 160 | 10 | Q8LKH4 forsythia 1  |
| 27 | 110   | 7.3 | 370 | 10 | Q9XEB0 arabidopsis  |
| 28 | 108.5 | 7.2 | 286 | 4  | Q8TER6 mus musculu  |
| 29 | 108   | 7.1 | 729 | 11 | Q8C7W3 mus musculu  |
| 30 | 107.5 | 7.1 | 288 | 16 | Q8E546 streptococc  |
| 31 | 107.5 | 7.1 | 305 | 17 | Q8TJ37 methanosarc  |
| 32 | 107.5 | 7.1 | 537 | 5  | Q8I470 plasmodium   |
| 33 | 107   | 7.1 | 925 | 11 | Q8C1I8 mus musculu  |
| 34 | 106   | 7.0 | 321 | 1  | Q5J138 methanosarc  |
| 35 | 106   | 7.0 | 327 | 17 | Q8PWJ3 methanosarc  |
| 36 | 105   | 6.9 | 288 | 2  | Q9F6G9 escherichia  |
| 37 | 104   | 6.9 | 318 | 16 | Q8DEP3 vibrio vuln  |
| 38 | 102.5 | 6.8 | 335 | 2  | Q8S0I1 methyllobact |
| 39 | 102   | 6.7 | 288 | 2  | Q9F6H0 escherichia  |
| 40 | 102   | 6.7 | 288 | 2  | Q9ETI6 escherichia  |
| 41 | 102   | 6.7 | 288 | 2  | Q9ETI2 escherichia  |
| 42 | 102   | 6.7 | 288 | 2  | Q9F6G7 escherichia  |
| 43 | 102   | 6.7 | 288 | 2  | Q9F6G0 escherichia  |
| 44 | 102   | 6.7 | 288 | 2  | Q9F6F4 escherichia  |
| 45 | 102   | 6.7 | 288 | 2  | Q9ETI5 escherichia  |

ALIGNMENTS

RESULT 1  
Q8EARD PRELIMINARY: PRT; 284 AA.  
ID Q8EARD; Q8EARD0;  
AC 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE 4-diphosphocytidylyl-2C-methyl-D-erythritol kinase.  
GN 1SPR OR S03836.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=22297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
RA Venter A., Weidman J., Imbraim N., Lee K., Berry K., Lee C.,  
RA Weller J., Kouri H., Gill J., Utterback T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
RT Shewanella oneidensis".  
RL Nat. Biotechnol. 20:1118-1123(2002).  
DR EMBL; AB015816; AAN56813.1; -.  
DR TIGR; S03836; -.  
KW Kinase; Complete proteome.  
SQ SEQUENCE 284 AA; 30721 MW; FE5A22B814242496 CRC64;

Query Match 42.5%; Score 644.5; DB 16; Length 284;  
Best Local Similarity 49.1%; Pred. No. 1.5e-52;  
Matches 135; Conservative 41; Mismatches 98; Indels 1; Gaps 1;

11 WPAPKXNLMIRITGRPDGYHLLQYVQKMLDCLDWLTFHVVDDGRVTLRPIFG-VPEQ 69

Db 9 WPAPAKLNLFLHNGRSDGYHELQTLFQVDCDDIDRVTDTPELILHSTKSAVADS 68  
 QY 70 DDLTVRAANLTKSHTCVAGVCIDIEKNLPMGGGLGSSSDAATTLVNLRLMGLSKRE 129  
 Db 69 DNLTRAAKSLQCATGFNGGAEIWLDRKLPWGGLGSSSDAATTLVNLRLMNTQLSHD 128  
 QY 130 ELMDLGLRLGADVPVFGCSANGEGVSEDLQATLTPGQWVFIKDCQVNIKGFSAFN 189  
 Db 129 ELAIGIKRLGADVPVFGHFAFGVGERLQAVNPAELMYLADPAHVSTAAVHODPL 188  
 QY 190 LITNSAVVTMSDFLAGNNDCEVWCXLYRFXKADIALLCYAEARLTGACVAPQPC 249  
 Db 189 LPRTNRLKGLDITLSCGFANDCGELVSKYPOVAKLGMILLEAPSRMGTGACVGERG 248  
 QY 250 NKEDASALEGLKDRMLVFLAKLINGSALYKKEQ 284  
 Db 249 SQOQALAAALAKPDSMDQGFVAKKMNISPLIVLNR 283

## RESULT 2

Q8DF6 PRELIMINARY; PRT; 290 AA.  
 AC Q8DF6;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate synthase.  
 GN V110256.  
 OS *Vibrio vulnificus*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxId=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Khee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
 RA Cho Y.H.E.;  
 RT "Complete genome sequence of *Vibrio vulnificus* CMCP6."  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB016797; AAC08792.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 290 AA; 31657 MW; 27161EC131ED8D CRC64;

Query Match 42.3%; Score 642; DB 16; Length 290;  
 Best Local Similarity 48.9%; Pred. No. 2,7e-52;  
 Matches 133; Conservative 43; Mismatches 96; Indels 0; Gaps 0;

QY 11 WPAPAKLNLMLRTGRPPDGYHLQTVFQMLDLCMLTFHPVDDGRVTLRNPISGVPEQ 70  
 Db 8 WSPAPAKLNLFLYNGRTDNGYHELQTLFQFLDHQDLTTANDSGHITLTPDIIVLPEQ 67  
 QY 71 DLTVRANLTKSHTCVAGVCIDIEKNLPMGGGLGSSSDAATTLVNLRLMGLSKRE 130  
 Db 68 NLTWKAANLQKRTGCTGATGANNILNKLPMGGGLGSSSDAATTLVNLRLMGLSKRE 127  
 QY 131 LMDLGLRLGADVPVFGCSANGEGVSEDLQATLTPGQWVFIKDCQVNIKGFSAFN 190  
 Db 128 LADIGIKRLGADVPVFGHFAFGVGERLQAVNPAELMYLADPAHVSTAAVHODPL 187  
 QY 191 LITNSAVVTMSDFLAGNNDCEVWCXLYRFXKADIALLCYAEARLTGACVAPQPC 249  
 Db 188 LPRTNRLKGLDITLSCGFANDCGELVSKYPOVAKLGMILLEAPSRMGTGACVGERG 247  
 QY 251 NKEDASALEGLKDRMLVFLAKLINGSALYKKE 282  
 Db 248 SQOQALAAALAKPDSMDQGFVAKKMNISPLIVLNR 279

RESULT 3  
 Q8F104 PRELIMINARY; PRT; 283 AA.  
 AC Q8F104;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148).  
 GN ISPE OR C1666.  
 OS *Escherichia coli* O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Escherichia*.  
 OX NCBI\_TaxId=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raebao D., Buckles E.L., Iden S.R., Boutin A., Hackett J., Stroud D.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 DR EMBL; AB016759; AAN80131.1; -  
 KW Kinase; Transferase; Complete proteome.  
 SQ SEQUENCE 283 AA; 30973 MW; E032729585EDDFC CRC64;

Query Match 40.9%; Score 620; DB 16; Length 283;  
 Best Local Similarity 47.9%; Pred. No. 3.1e-50;  
 Matches 135; Conservative 40; Mismatches 95; Indels 12; Gaps 4;

QY 10 WPAPAKLNLMLRTGRPPDGYHLQTVFQMLDLCMLTFHPVDDGRVTLRNPISGVPEQ 69  
 Db 4 QMPSPAKLNLFLYTGQADYHQLQTLFQFLDYGDITISIRLDGDIRLLTPVGVHE 63  
 QY 70 DDLTVRAANLTKSHTCVAGVCIDIEKNLPMGGGLGSSSDAATTLVNLRL 121  
 Db 64 DNLTRAAKSLQCATGFNGGAEIWLDRKLPWGGLGSSSDAATTLVNLRL 121  
 QY 122 WGLSKRELMDLGLRLGADVPVFGCSANGEGVSEDLQATLTPGQWVFIKDCQVNI 181  
 Db 122 WQCGISMDLGLRLGADVPVFGHFAFGVGERLQAVNPAELMYLADPAHVSTAAVHODPL 181  
 QY 182 GEIPSAENLITNSAVVTMSDFLAGNNDCEVWCXLYRFXKADIALLCYAEARLTGAC 241  
 Db 182 PVTFQDPELPRTNRLKGLDITLSCGFANDCGELVSKYPOVAKLGMILLEAPSRMGTGAC 241  
 QY 242 ACVFAPOCNKEDASALEGLKDRMLVFLAKLINGSALYKKE 282  
 Db 242 ACVFAFPDTESEARQVLEQADE-WLNGFVAKGVNLSPLIRAM 282

RESULT 4  
 Q8D2K6 PRELIMINARY; PRT; 304 AA.  
 AC Q8D2K6;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE YcbB protein.  
 GN YCB.  
 OS *Mycobacterium tuberculosis* H37Rv.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Mycobacterium*.  
 OX NCBI\_TaxId=164609;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22297718; PubMed=12219091;  
 RA Akman U., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 RA Aksoy S.;  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 fly, *Mycobacterium tuberculosis* H37Rv."  
 RL Nat. Genet. 32:402-407 (2002).  
 DR EMBL; AB063522; BAC24494.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 304 AA; 34682 MW; FB746B5858189985 CRC64;



Query Match 33.9%; Score 513.5; DB 16; Length 304;  
 Best Local Similarity 39.3%; Pred. No. 3.9e-40;  
 Matches 110; Conservative 63; Mismatches 98; Indels 9; Gaps 2;

QY 11 WPAPAKINIMLRITGRRRPDGTHLQTFVQMLDLCWMLTFHPVDDGRVTLNRPISGVPEOD 70  
 DB 18 WEPAPAKINIMLRITGRRRPDGTHLQTFVQMLDLCWMLTFHPVDDGRVTLNRPISGVPEOD 77  
 QY 71 DLTVRAANLI-----KSHTCVGRVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMG 123  
 DB 78 NLIRRAKSKDKPMWKKKND--TPGVKFIKKYIPIGGGLGGSSNAAATTLALNHNK 135  
 QY 124 LGSKREIMDGLRIADGVFVFGSAGMEGSEDDQATLPEQMPVLIKPDCHVNTB 183  
 DB 136 CKLSLNTLADYGLQIGIDIVFYGKSAFAGGKLSLFPQPKKYLIVIPKISTKL 195  
 QY 184 IFSANLTRNSAVVMSDFLAGNRNDCSEVCKLYRPVMDAIDALLCYAENLTGTAC 243  
 DB 196 IFNKEFLNKSKYLSCKQYIKKKPLKNDPBPWKRNFIIDNLINYSKSNFRLTGTGSC 255  
 QY 244 VFAQPCNKEDASALBGLKRWLVFLAKGLNOSALTKKE 283  
 DB 256 IFSEFSECKAKKILYKLPKIKGFWKGTINISYLKELIK 295

RESULT 5  
 OEB7K5 PRELIMINARY; PRT; 283 AA.

AC OEB7K5;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Hypothetical protein.  
 GN GBS0149.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaeser P., Rasmok C., Buchrieser C., Chevallier F., Frangeul L.,  
 RA Maeder T., Zouine M., Couve E., Lailouli E., Poyart C., Tillet-Cuot P.,  
 RA Kunst F.;  
 RT "genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RT Mol. Microbiol. 45:1489-1513 (2002).  
 DR EMBL; AL766844; CAD45794.1; -;  
 DR Sagalst; GBS0149; -;  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 283 AA; 31075 MW; D195DA8A84CC0FE0 CRC64;

Query Match 21.9%; Score 332; DB 16; Length 283;  
 Best Local Similarity 32.2%; Pred. No. 4.8e-23;  
 Matches 85; Conservative 58; Mismatches 107; Indels 14; Gaps 5;

QY 13 APAKINIMLRITGRRRPDGTHLQTFVQMLDLCWMLTFHPVDDGRVTLNRPISGVPEOD 72  
 DB 7 APAKINIMLRITGRRRPDGTHLQTFVQMLDLCWMLTFHPVDDGRVTLNRPISGVPEOD 66  
 QY 73 TV-RAANLKSHTGCVRGVCIIDIEKNLPMGGGLGGSSDAATTLVNLRLMGISKREL 131  
 DB 67 DVFKADITIKNGYKNGVIRLKSIPVCAIGGSDTAATIRALNRLNLMQMDYDEM 126  
 QY 132 MDGLRLGADVPVFGSAGMEGSEDDQATLPEQMPVLIKPDCHVNTGEIF--SA 187  
 DB 127 VAIGFKISDVPYGLGGGSLVKGELIVKPLTPKPCWIVLVKPDGISTKSIREDIDC 186  
 QY 188 ENLTR-----NSAVVMSDFLAGNRNDCSEVCKLYRPVMDAIDALLCYA---EARLT 238  
 DB 187 KSISRVDIDLKSAISDYQLMVKWSKNSLEDITTKKPVISTIKERLNSGADVAMT 246

QY 239 GTGACVFAQPCNKEDASALBGLK 262  
 DB 247 GSGPTVFMCKSTKKAADRVFNSMK 270

RESULT 6  
 OEB245 PRELIMINARY; PRT; 283 AA.

AC OEB245;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE 4-diphosphocyclidyl-2C-methyl-D-erythritol kinase.  
 GN ISP OR SAG0153  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=2222988; PubMed=12200547;  
 RA Tettelin H., Mastignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
 RA Radune D., Fedorova N.B., Scanlan D., Kouri H., Mulligan S.,  
 RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
 RA Iacobini E.T., Breton C., Galli G., Mariotti M., Vignati F., Malone D.,  
 RA Rinaldo D., Rappelli R., Telford J.L., Kasper D.L., Grand G.,  
 RA Frazer C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).  
 DR EMBL; AB014198; AA09061.1; -;  
 DR TIGR; SAG0153; -;  
 KM Kinase; Complete proteome.  
 SQ SEQUENCE 283 AA; 31047 MW; FC93843F69C2DC0D CRC64;

Query Match 21.9%; Score 332; DB 16; Length 283;  
 Best Local Similarity 32.2%; Pred. No. 4.8e-23;  
 Matches 85; Conservative 58; Mismatches 107; Indels 14; Gaps 5;

QY 13 APAKINIMLRITGRRRPDGTHLQTFVQMLDLCWMLTFHPVDDGRVTLNRPISGVPEOD 72  
 DB 7 APAKINIMLRITGRRRPDGTHLQTFVQMLDLCWMLTFHPVDDGRVTLNRPISGVPEOD 66  
 QY 73 TV-RAANLKSHTGCVRGVCIIDIEKNLPMGGGLGGSSDAATTLVNLRLMGISKREL 131  
 DB 67 DVFKADITIKNGYKNGVIRLKSIPVCAIGGSDTAATIRALNRLNLMQMDYDEM 126  
 QY 132 MDGLRLGADVPVFGSAGMEGSEDDQATLPEQMPVLIKPDCHVNTGEIF--SA 187  
 DB 127 VAIGFKISDVPYGLGGGSLVKGELIVKPLTPKPCWIVLVKPDGISTKSIREDIDC 186  
 QY 188 ENLTR-----NSAVVMSDFLAGNRNDCSEVCKLYRPVMDAIDALLCYA---EARLT 238  
 DB 187 KSISRVDIDLKSAISDYQLMVKWSKNSLEDITTKKPVISTIKERLNSGADVAMT 246

RESULT 7  
 OEB540 PRELIMINARY; PRT; 282 AA.

AC OEB540;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Putative isopenicillin monophosphate kinase.  
 GN IPK OR SMU.1996.

OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 CX NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
 RX MEDLINE=22295063; PubMed=12397186;  
 RA Ajdic D., Mcshan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.D.;  
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
 pathogen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 DR EMBL: AE015023; AAN59600.1;  
 KW Kinase; Complete proteome.  
 SO SEQUENCE 282 AA; 30964 MW; 65FD3629BF884289 CRC64;

Query Match 21.2%; Score 321; DB 16; Length 282;  
 Best Local Similarity 32.5%; Pred. No. 5,2e-22;  
 Matches 87; Conservative 57; Mismatches 102; Indels 22; Gaps 9;

QY 13 APATLMMLRTGRPPGTHLQTVFQMDLCPMLTFHPVDDGRVTLRNPISGV-EGD 70  
 DB 7 APAKINIGLDLGGKYQGGFHELSKIMASVDLNDYLTITTEIAEKIVESNNCKLPNRKN 66  
 QY 71 DLTVRAANILKSHRTGCVRGVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSKRE 130  
 DB 67 DV-YKAAHLKRRHHISTGDKISLQKKIPICAGIGSSSDPAATLRALNCLMKNLSPE 125  
 QY 131 LMDIGRLGADVVFV-FGGS-AWGEVSBDLQATILPEQWFTIKKDCVNTGEIFSAE 188  
 DB 126 LIDPFGISGDPVPCIEAGCALISGKGIIVEPLATTL-STWVVLVKDFPISITTFKEI 184  
 QY 189 NLTNSAV-----VTMSDF-----LAGDNRDCESEVCKYRPVKAIDALLCYAE- 234  
 DB 185 DMATISVDLPALKEALIANYEALQPMNSLEDITAKKPTQIKGM-IKGADI 242  
 QY 235 ARLTGTACVFAQPCNKEDAESALGLK 262  
 DB 243 ALMTGSGTVEALCRTERADRVVSMK 270

RESULT 8  
 Q8G2D0 PRELIMINARY; PRT; 299 AA.  
 ID Q8G2D0  
 AC Q8G2D0;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DB 4-diphosphocytidylyl-2C-methyl-D-erythritol kinase.  
 GN ISPE OR BR0394.  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 CX NCBI\_TaxID=29461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22247741; PubMed=12271122;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,  
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.B.,  
 RA Redmiller S., Tellezn H., Gill S.R., White O., Salzberg S.L.,  
 RA Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;  
 RT "The Brucella suis genome reveals fundamental similarities between  
 animal and plant pathogens and symbionts."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL: AE014350; AAN29340.1;  
 KW TIGR: BR0394;  
 KM Kinase; Complete proteome.  
 SO SEQUENCE 299 AA; 31188 MW; A9B88F4157BF513 CRC64;

Query Match 20.7%; Score 313.5; DB 16; Length 299;  
 Best Local Similarity 34.5%; Pred. No. 2.9e-21;  
 Matches 92; Conservative 42; Mismatches 104; Indels 29; Gaps 11;

QY 13 APATLMMLRTGRPPGTHLQTVFQMDLCPMLTFHPVDDGRVTLRNPISGV-EGD 70  
 DB 15 APAKINIGLDLGGKYQGGFHELSKIMASVDLNDYLTITTEIAEKIVESNNCKLPNRKN 66  
 QY 71 DLTVRAANILKSHRTGCVRGVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSKRE 129  
 DB 75 NIVLKRDLRQHGFDLSPVAIHLEKNLPIASIGIGSSDAATTLALNTLMQDLDFE 134  
 QY 130 ELMDDIGRLGADVVFVFGCSAMG-----EGVSDLQATILPEQWFTIKKDCVNTGEIFSAE 180  
 DB 135 MIAATIGLIGADIPKCIHG-AAHGTPILARKIGBELNDVSGIAALP--MLVNDGTALA 190  
 QY 181 TGEIFSA---ENLTNSAVVTMSDFLAG--DNRDCSEVCKYRPVKAIDALLCYA 233  
 DB 191 TEDVFFALTRRRNAPLPPACGTDLCAYLRETRNDLLPAISLAPQIEPKALIRAKG 250  
 QY 234 E--ARLTGTACVFAQPCNKEDAESAL 258  
 DB 251 ALYQMSGGATCFALFSD---ESAL 273

RESULT 9  
 Q8C0U6 PRELIMINARY; PRT; 282 AA.  
 ID Q8C0U6  
 AC Q8C0U6;  
 DT 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DB 4-diphosphocytidylyl-2C-methyl-D-erythritol kinase.  
 GN SE2288.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CX NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AE016751; AA005930.1;  
 KW Kinase; Complete proteome.  
 SO SEQUENCE 282 AA; 31030 MW; ASD58B0A01D48287 CRC64;

Query Match 19.9%; Score 302; DB 16; Length 282;  
 Best Local Similarity 32.0%; Pred. No. 3.2e-20;  
 Matches 79; Conservative 51; Mismatches 103; Indels 14; Gaps 6;

QY 13 APATLMMLRTGRPPGTHLQTVFQMDLCPMLTFHPVDDGRVTLRNPISGV-EGD 71  
 DB 6 APAKINIGLDLGGKYQGGFHELSKIMASVDLNDYLTITTEIAEKIVESNNCKLPNRKN 65  
 QY 72 DLTVRAANILKSHRTGCVRGVCIDIEKNLPMGGGLGGSSDAATTLVNLRLMGLSKREL 131  
 DB 66 LAYKADLFFERFENINBGTISIDKIPASALAGSADAAATMGELNRLFELGSLDAL 125  
 QY 133 MDIGRLGADVVFVFGCSAMGEGVSDLQATILPEQWFTIKKDCVNTGEIFSAE 190  
 DB 126 ALGIGIGIDIPFCIYNGTAVCTGGEVTFPKRPPSAMVVLAKENIGISSDVFKADEL 185  
 QY 191 TNSAV-----VTMSDF--LAQDNRDCESEVCKYRPVKAIDALLCYA-ARLT 238  
 DB 186 TEHIVHEKCKQALENNHYLLCNLSLNRLPEVSMAMPDIKKIKDNNMLGAGDALMS 245  
 QY 239 GTGACVF 245  
 DB 246 GSGPTVY 252



DB 12 LQWSEIGSDIPFFFSQAAVCTGRGB-IVEDIHNPANLP---MVLKPEACSTAE 267

QY 184 IFS---AENLTRNSAVYTMDELPLAGNRNDSCSEVVC 216

DB 268 VYRRLRLRHTSQTDPVLKELI---TENGISQDAC 299

RESULT 13

Q86FQ24 PRELIMINARY; PRT; 315 AA.

AC Q86FQ24; 01-MAR-2003 (TREMblrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE Putative isopentenyl-monophosphate kinase.

GN CE0973.

OS Corynebacterium efficiens.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.

OX NCBI\_TaxID=152794;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;

RA Kawabata Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y., Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y., Ueda Y., Sugimoto S.;

RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";

RT Submitted (MAY-D002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AP005217; BAC17783.1; -

KM Kinase; Complete proteome.

SO SEQUENCE 315 AA; 32871 MW; 472052CC7D843763 CRC64;

Query Match 14.7%; Score 223.5; DB 16; Length 315;

Best Local Similarity 28.3%; Pred. No. 9.6e-13;

Matches 80; Conservative 46; Mismatches 116; Indels 41; Gaps 12;

QY 4 AAGGGEWPAKALNEMLRITGRRPDGYHLLQTVFQMLDCMLTFHPVD----- 54

DB 5 ARAGW-----KINMLGVGPARDEGHVLFQITDLAETITLTLEDELVEEGSV 57

QY 55 GRVTLRPISGV-EQDDLTVRANLPL---KSHTGC-VRGVCIDIEKNLPMGGGLGGGS 108

DB 58 RQILVTRP-RGVPTTPNLMRAVDALVGRRRHEDRPPLAVLHIDKGLPVGAGAGGS 116

QY 109 SDAATTLVLRNLGLSKRELMDLGLRLGADVPVVFVFGCSAMGEGVSEDLQATLP--P 166

DB 117 ADAAAALRAVD-AWIGPFGETILEVAALGSDVPFCLGRTLGTGRGQLVDMLSRGT 175

QY 167 EQMFTIIRPDCHVNVGSIIFS---AENLTRNSAVYTMDELPLAGNRNDSCSEVVC 222

DB 176 YHMLLVSPK-GISTPEVFAKFPDMSLPSSMDVTPMSQALLDSAGALAEVLENDLAPAA 234

QY 223 -----KDAIDALLCYAERLITGTGACVPAQFCNKEDAES 257

DB 235 LSLRPDLRKTQLAGIRAGALATWVGSGPTIALLC--DDAQSA 275

RESULT 14

Q86G14 PRELIMINARY; PRT; 316 AA.

AC Q86G14; 01-MAR-2003 (TREMblrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase, Cmk, 4 (Cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase.

GN ISPP OR BL0656.

OS Bifidobacterium longum.

OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;

OC Bifidobacteriaceae; Bifidobacterium.

OX NCBI\_TaxID=216816;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCC 2705;

RX MEDLINE=22294977; PubMed=12381787;

RA Schell M.A., Karmaliatou M., Snel B., Vilanova D., Berger B., RA Pestl G., Zwielen M.-C., Desiere F., Bork P., Delley M., RA Fridmore R.D., Arigoni F.;

RT "The genome sequence of Bifidobacterium longum reflects its adaptation to the human gastrointestinal tract";

RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).

DR EMBL: AE014687; AAN24478.1; -

KM Kinase; Complete proteome.

SO SEQUENCE 316 AA; 33286 MW; AEF623697DB7875B CRC64;

Query Match 13.5%; Score 204.5; DB 16; Length 316;

Best Local Similarity 25.8%; Pred. No. 6e-11;

Matches 74; Conservative 44; Mismatches 118; Indels 51; Gaps 8;

QY 14 PAKMLMRLITGRPD--GYHLLQTVFQMLDCMLTFHPVDGRTLTNPISGVPEQ-- 69

DB 30 PAKMLTLEVGPAHDEWGRHEDLTIYCAIGYDTVT-----ATAKQPAQFSLE 81

QY 70 ----DDLTVRANLTKSHRGV-----RGVCIDIEKNLPMGGGLGGSSDAAT 114

DB 82 GAYLDLSSRSRDMRANALALPMAQAEREPVALITTKRIFVGAIGGGSDAAT 141

QY 115 LVVLRNLGLSKRELMDLGLRLGADVPVVFVFGCSAMGEGVSEDL-----QATLP 166

DB 142 MLAVRMLWELMPLIRLTIAATLADMPFCLTGLAVGTGFRITDIAPSRDELALI 201

QY 167 EQMFTIIRPDCHVNVGSIIFS---AENLTRNSAVYTMSPF-----LAGNRNDSCSEVVC 222

DB 202 EQGF-----GSEVLVGVYOSQSTPEYHTFTDIVGAEGGRNHLQAAALSLHPS 251

QY 223 KDAIDALLCYAERLITGTGACVPAQFCNKEDAESLEGLDRMLV 267

DB 255 GQAIIDAAQAGASHAFVSGSGPSVAFADAAARIIIEWRDTAVV 298

RESULT 15

Q8EZM8 PRELIMINARY; PRT; 297 AA.

AC Q8EZM8; 01-MAR-2003 (TREMblrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE 4-diphosphocytidylyl-2C-methyl-D-erythritol kinase (EC 2.7.1.-).

GN YCHB OR LA3824.

OS Leptospira interrogans.

OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.

OX NCBI\_TaxID=173;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;

RA Ren S.;

RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AE011537; AAN51022.1; -

KM Kinase; Transferase; Complete proteome.

SO SEQUENCE 297 AA; 33271 MW; 51A1F018DAA7872 CRC64;

Query Match 12.9%; Score 196; DB 16; Length 297;

Best Local Similarity 26.3%; Pred. No. 3.5e-10;

Matches 77; Conservative 53; Mismatches 115; Indels 48; Gaps 11;

QY 13 AAKNLMRLITGRPDGYHLLQTVFQMLDCMLTFHPVDGRTLT-----RNP 62

DB 3 SPANINLGLIEFPKSLDGFHRSVFLKISMDDIRIPASGVVELSNNEILLKRL 62

QY 63 ISGVEQDILT-----VRAANLTKSHTGCVRVCIDIEKNLPMGGGLGGSSDAAT 114

DB 63 YQVSEGRGDKINNIYKTFIKARSLPR---LPVKIHLTRIPAGAGLGGSTMAAS- 117

QY 115 LVVLRNLGLSKRELMDLGLRLGADVPVVFVFGCSAMGEGVSEDLQATLP--QMFV 171

DB 118 --LNFPLSKMRPFSTSDMFVLAETIGSDVPFCLGRTLGTGRGQLVDMLSRGT 175

|    |     |   |     |
|----|-----|---|-----|
| QY | 172 | IIPDCHVTGELIS-----ABNLTRN-SAVVTMSDF--LAGDRNRDCSEV         | 214 |
| Db | 176 | ALTPQV-MNTSEMYSLIKPLQESASOKNGNTLSKNLISILKNGDWSLQGLRNDPEEV | 234 |
| QY | 215 | VCKLYRPVKDAIDALLCYAEA--RLNGTACVFAQFCNKEDAESALEGLKDRW      | 265 |
| Db | 235 | AFOLHPBELAVLKDKFLFEGSSYCSLNGSGSSMYGLVQGLEIQEELLQRLRQEF    | 287 |

Search completed: January 29, 2004, 15:54:08  
 Job time : 23.2436 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 15:40:33 : Search time 14.3511 Seconds  
(without alignments)  
1736.455 Million cell updates/sec

Title: US-09-941-947a-14

Perfect score: 806

Sequence: 1 MIRXGCGYDHRFNDGDHIL.....LGFGRKRGIAVAVVLIEN 157

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A\_Geneseq\_19Jun03.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 806   | 100.0       | 157    | 23    | ABG61583    |
| 2          | 806   | 100.0       | 157    | 23    | AAAB22305   |
| 3          | 806   | 100.0       | 157    | 23    | AAAB22305   |
| 4          | 537   | 66.6        | 159    | 21    | AAAB11373   |
| 5          | 537   | 66.6        | 159    | 21    | AAAB27149   |
| 6          | 537   | 66.6        | 159    | 22    | AAAG38111   |
| 7          | 537   | 66.6        | 159    | 24    | ABP70685    |
| 8          | 534   | 66.3        | 158    | 21    | AAAB7150    |
| 9          | 534   | 66.3        | 158    | 21    | AAAB7151    |

|    |       |      |     |    |           |
|----|-------|------|-----|----|-----------|
| 10 | 531   | 65.9 | 159 | 24 | ABP70686  |
| 11 | 528   | 65.5 | 159 | 24 | ABP70687  |
| 12 | 528   | 65.5 | 159 | 24 | ABP70688  |
| 13 | 519   | 64.4 | 170 | 24 | ABG75686  |
| 14 | 507   | 62.9 | 174 | 24 | ABP79612  |
| 15 | 505   | 62.7 | 160 | 24 | ABU06061  |
| 16 | 451   | 56.0 | 157 | 23 | ABG49155  |
| 17 | 438   | 54.3 | 191 | 22 | ABG18263  |
| 18 | 405   | 50.2 | 120 | 23 | ABP02539  |
| 19 | 396   | 49.1 | 161 | 21 | AAAB27152 |
| 20 | 377   | 46.8 | 231 | 21 | AAAB15276 |
| 21 | 377   | 46.8 | 240 | 21 | AAAB15275 |
| 22 | 374   | 46.4 | 231 | 21 | AAAG27449 |
| 23 | 374   | 46.4 | 239 | 21 | AAAG27448 |
| 24 | 353   | 43.8 | 259 | 21 | AAAG27633 |
| 25 | 345   | 42.8 | 409 | 19 | AAAB1013  |
| 26 | 343   | 42.6 | 406 | 21 | AAAB27154 |
| 27 | 341   | 42.3 | 352 | 18 | AAW20628  |
| 28 | 329   | 40.8 | 399 | 21 | AAAB27153 |
| 29 | 321   | 39.8 | 223 | 23 | ABP93990  |
| 30 | 286   | 35.5 | 176 | 22 | AAU50581  |
| 31 | 269   | 33.4 | 240 | 21 | AAAB18226 |
| 32 | 269   | 33.4 | 248 | 22 | AAAB70488 |
| 33 | 247.5 | 30.7 | 158 | 24 | AAE31686  |
| 34 | 242   | 30.0 | 161 | 21 | AAAG39032 |
| 35 | 229   | 28.4 | 174 | 23 | ABP65843  |
| 36 | 213.5 | 26.5 | 160 | 22 | AAAG2667  |
| 37 | 213.5 | 26.5 | 160 | 22 | AAAB0182  |
| 38 | 153   | 19.0 | 199 | 20 | AAAB35167 |
| 39 | 151   | 18.7 | 96  | 18 | AAAB20173 |
| 40 | 117.5 | 14.6 | 102 | 20 | AAAB37797 |
| 41 | 86    | 10.7 | 434 | 21 | AAAG08216 |
| 42 | 83.5  | 10.4 | 410 | 23 | ABP28012  |
| 43 | 82    | 10.2 | 943 | 22 | AAAG91445 |
| 44 | 81    | 10.0 | 434 | 21 | AAAG38717 |
| 45 | 81    | 10.0 | 434 | 23 | ABP92639  |

#### ALIGNMENTS

|          |   |   |
|----------|---|---|
| RESULT 1 | ABG61583  | High growth methanobacterial strain polypeptide #3. |
| ID       | ABG61583  | standard; Protein; 157 AA.                          |
| AC       | ABG61583  |   |
| XX       | XX  |   |
| DT       | 27-AUG-2002 (first entry)   |   |
| XX       | XX  |   |
| DE       | High growth methanotrophic bacterial strain polypeptide #3.   |   |
| XX       | XX  |   |
| KM       | High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme; methane; methanol; Embden-Meyerhof carbon flux pathway; 16S rRNA;     |   |
| KM       | pyrophosphate dependent phosphofructokinase; nitrogen-containing compound; ammonium; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;     |   |
| KM       | methane-containing environment; waste water treatment system; isoprenoid; nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide. |   |
| XX       | XX  |   |
| OS       | Methylobionas 16a.  |   |
| XX       | XX  |   |
| FN       | W020020728-A2.  |   |
| XX       | XX  |   |
| PD       | 14-MAR-2002.  |   |
| XX       | XX  |   |
| PF       | 28-AUG-2001; 2001WO-US26827.  |   |
| XX       | XX  |   |
| PR       | 01-SEP-2000; 2000US-229858P.  |   |
| XX       | XX  |   |
| PA       | (DUPO) DU PONT DE NEMOURS & CO E. I.  |   |
| XX       | XX  |   |
| PI       | Koffas M, Odem JM, Schenzle A;  |   |
| XX       | XX  |   |
| DR       | WPI; 2002-452200/48.  |   |

2C-methyl-D-erythr  
2C-methyl-D-erythr  
2C-methyl-D-erythr  
2C-methyl-D-erythr  
N. gonorrhoeae and  
N. meningitidis va  
Listeria monocyco  
Novel human diagno  
Human ORF protein  
Synchocystis esse  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
H. pylori ORF hp7e  
H. pylori ORF hp7e  
H. pylori cytoplasm  
T. pallidum essent  
Herbicidally activ  
Propionibacterium  
Plasmodium falcipar  
Plasmodium falcipar  
Rhodococcus erythr  
Arabidopsis thalia  
Bifidobacterium 10  
C glutamicum prote  
Chryseobacterium 91  
H. pylori cytoplasm  
H. pylori cytoplasm  
Amino acid sequenc  
Arabidopsis thalia  
Streptococcus poly  
C glutamicum prote  
Arabidopsis thalia  
Herbicidally activ

DR N-PSDB; ABR63262.

XX New high growth methanotrophic bacterial strain, useful for producing  
 PT single cell proteins, grows on a C1 carbon substrate, and comprises a  
 PT functional gene encoding in Embden-Meyerhof carbon pathway -

PS Claim 11; Page 142-143; 157pp; English.

CC The invention relates to a high growth methanotrophic bacterial strain,  
 CC which grows on a C1 carbon substrate e.g. methane and methanol, and  
 CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a  
 CC gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16s  
 CC rRNA. The bacterial strain is useful for the production of single cell  
 CC protein and for the biotransformation of a nitrogen-containing compound,  
 CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the  
 CC production of a feed product comprising a protein, carbohydrates and a  
 CC pigment and for reducing oxygen demand, for removing nitrates and  
 CC nitrates in methane-containing environments such as landfills, waste  
 CC water treatment systems or anywhere that methane, oxygen and nitrates are  
 CC present. The bacterial strain of the invention can be used as a  
 CC denitrifying agent for the conversion of nitrate or nitrite to nitrous  
 CC oxide with methane or methanol as a carbon source. It is also used in the  
 CC production of biomass including proteins, carbohydrates and a wide  
 CC variety of pigments (particularly for isoprenoid pigments for the  
 CC purpose of generating animal feeds), in production of terpenoid and  
 CC carotenoid compounds, useful as pigments and as monomers in polymeric  
 CC materials and in production of exopolysaccharides at high levels.  
 CC Sequences ABG61551-ABG61590 represent high growth methanotrophic  
 CC bacterial strain proteins of the invention.

SO Sequence 157 AA;

Query Match 100.0%; Score 806; DB 23; Length 157;

Best Local Similarity 100.0%; Pred. No. 1.2e-87;

Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60  
 DB 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60  
 QY HPPDTPNFKGASRVLLRHVYGVIVKKGKYLNAVDTTIIAQAQKMLPHVPGKRNIAAD 120  
 DB HPPDTPNFKGASRVLLRHVYGVIVKKGKYLNAVDTTIIAQAQKMLPHVPGKRNIAAD 120  
 QY 121 LETDVDFINVKATTTTEKLGEGRKGIAVQAVVLIIR 157  
 DB 121 LETDVDFINVKATTTTEKLGEGRKGIAVQAVVLIIR 157

RESULT 2

AAE22305 standard; Protein; 157 AA.

AC AAE22305;

DT 25-JUL-2002 (first entry)

DE Methylomonas 16a 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase.

KW Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;  
 KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;  
 KW aquaculture; 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase; ispF;  
 KW enzyme.

OS Methylomonas 16a sp.

PN WO200218617-A2.

PD 07-MAR-2002.

PF 04-SEP-2001; 2001MO-US27420.

PR 01-SEP-2000; 2000US-229856P.

PR 01-SEP-2000; 2000US-229907P.

PA (DUPO) DU PONT DE NEMOURS &amp; CO B I.

PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

PI Odom JM, Picataggio SK, Rouviere PB;

DR WPI; 2002-351711/38.

DR N-PSDB; AAD35503.

PT Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by  
 PT using microorganisms having a nucleic acid molecule encoding enzymes in  
 PT the carotenoid biosynthetic pathway and which metabolize single carbon  
 PT substrates -

PS Claim 45; Page 118-119; 156pp; English.

CC The invention relates to a method for producing carotenoid compounds.  
 CC The method comprises a transformed metabolizing host cell, comprising  
 CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule  
 CC encoding an enzyme in the carotenoid biosynthetic pathway, under the  
 CC control of regulatory sequences, and contacting the host cell with carbon  
 CC substrate to produce a carotenoid compound. The method is useful for  
 CC producing carotenoid compounds such as antheraxanthin and astaxanthin, by  
 CC using microorganism having a nucleic acid molecule encoding enzymes in  
 CC the carotenoid biosynthetic pathway and which metabolize single carbon  
 CC substrates. The carotenoid have potent anti-oxidant properties useful in  
 CC diet, and aquaculture elements. The carotenoids are also useful as  
 CC intermediates in the synthesis of steroids flavours and fragrances and  
 CC compounds for potential electro-optic applications. The present sequence  
 CC is Methylomonas 16a sp. 2C-methyl-d-erythritol 2,4-cyclodiphosphate  
 CC synthase (ispF) enzyme used in the invention.

SQ Sequence 157 AA;

Query Match 100.0%; Score 806; DB 23; Length 157;

Best Local Similarity 100.0%; Pred. No. 1.2e-87;

Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60  
 DB 1 MIRVGAGYVHRNDGDHIIIGGVKIPYKGLFAHSDGVVLAALADALIGAAALGDIGK 60  
 QY HPPDTPNFKGASRVLLRHVYGVIVKKGKYLNAVDTTIIAQAQKMLPHVPGKRNIAAD 120  
 DB HPPDTPNFKGASRVLLRHVYGVIVKKGKYLNAVDTTIIAQAQKMLPHVPGKRNIAAD 120  
 QY 121 LETDVDFINVKATTTTEKLGEGRKGIAVQAVVLIIR 157  
 DB 121 LETDVDFINVKATTTTEKLGEGRKGIAVQAVVLIIR 157

RESULT 3

AAU80329 standard; Protein; 157 AA.

AC AAU80329;

DT 15-JUL-2002 (first entry)

DE Methylomonas 16a ORF5 ygbB/ispF protein sequence.

KW Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
 KW keratinoid; pigment; flavour; fragrance; open reading frame 5; ORF5;  
 KW ygbB; 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase enzyme; ispF.

OS Methylomonas sp.

PN WO200220733-A2.

PD 14-MAR-2002.

PR 29-AUG-2001; 2001MO-US26852.





XX (GPCB-) GPC BIOTECH AG.  
 PA Loferer H, Jacobi A;  
 PI MPI; 2000-687048/67.  
 DR  
 XX Identifying antibacterial compounds, comprises identifying an  
 PT antagonist or inhibitor of the expression of a gene encoding a  
 PT polypeptide essential for bacterial growth or survival -  
 PS Disclosure; Fig 4; 75pp; English.  
 XX  
 CC The present invention relates to antagonists and inhibitors of 24  
 CC bacterial genes and proteins. The proteins are thought to be essential  
 CC for growth in several species of bacteria (including *S. pneumoniae*, *B.*  
 CC *burgorferi*, *H. influenza* and *H. pylori*). The proteins and coding  
 CC sequences shown in the specification can be used to identify antagonists  
 CC and inhibitors which can be used in disease treatment and pesticides. In  
 CC particular, they can be used against *M. tuberculosis*. The present  
 CC sequence is one of the proteins of the invention.  
 CC  
 SQ Sequence 159 AA;  
 Query Match 66.6%; Score 537; DB 21; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 1.4e-55;  
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;  
 QY 2 IRVGMGYDVERFENDGHIILGGVYKIPYKGLIAHSDGSDVLAHALADALIGAAALGDIKX 61  
 DB 1 MRIGHGFVHAFGGEGPILIGVRIPIYKGLIAHSDGSDVLAHALADALIGAAALGDIKX 60  
 QY 62 FPDIDPNEFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVHVGKRAMIAADL 121  
 DB 61 FPDIDPNEFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVHVGKRAMIAADL 120  
 QY 122 ETDVDFINVKATTTKLGFBGRKEGIAVQAVLI 155  
 DB 121 GCHMDVNVKATTTKLGFBGRKEGIAVQAVLI 154  
 RESULT 6  
 AAG63811  
 ID AAG63811 standard; Protein; 159 AA.  
 AC AAG63811;  
 XX  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase.  
 XX  
 KM 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase;  
 KM non-mevalonate pathway; 2-C-methyl-D-erythritol-2,4-cyclodiphosphate;  
 KM 2-phospho-4-(cytidine 5'-diphospho) 2-C-methyl-D-erythritol; CDP-ME2P;  
 KM cytidine 5'-monophosphate; isoprenoid; antibacterial.  
 XX  
 OS *Escherichia coli*.  
 XX  
 PN WO200157223-A1.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 25-JAN-2001; 2001W0-JP00483.  
 XX  
 PR 03-FEB-2000; 2000US-0180126.  
 PR 07-FEB-2000; 2000JP-0029287.  
 XX  
 PA (SETO/) SETO H.  
 PA (KUZUYAMA T. KUZUYAMA T.  
 XX  
 PI Seto H, Kuzuyama T;  
 XX MPI; 2001-522353/57.  
 DR

DR N-PSDB; AAR74788.  
 XX  
 PT 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase, useful for  
 PT producing isoprenoids and for screening for non-mevalonate pathway  
 PT inhibitors -  
 XX  
 PS Claim 2; Page 42-43; 49pp; Japanese.  
 XX  
 CC The present sequence represents a  
 CC 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthetase. The enzyme is  
 CC involved in the non-mevalonate pathway. It acts on  
 CC 2-phospho-4-(cytidine 5'-diphospho) 2-C-methyl-D-erythritol (CDP-ME2P) -  
 CC to produce 2-C-methyl-D-erythritol-2,4-cyclodiphosphate and cytidine  
 CC 5'-monophosphate. The enzyme has a molecular weight of 22 kDa, as  
 CC determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis.  
 CC The enzyme or gene can be used to increase isoprenoid compound  
 CC production, and also for screening for antibacterials.  
 CC  
 SQ Sequence 159 AA;  
 Query Match 66.6%; Score 537; DB 22; Length 159;  
 Best Local Similarity 69.5%; Pred. No. 1.4e-55;  
 Matches 107; Conservative 14; Mismatches 33; Indels 0; Gaps 0;  
 QY 2 IRVGMGYDVERFENDGHIILGGVYKIPYKGLIAHSDGSDVLAHALADALIGAAALGDIKX 61  
 DB 1 MRIGHGFVHAFGGEGPILIGVRIPIYKGLIAHSDGSDVLAHALADALIGAAALGDIKX 60  
 QY 62 FPDIDPNEFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVHVGKRAMIAADL 121  
 DB 61 FPDIDPNEFKGADSRVLLRHVYGIYKGYKLVNADVTIIIAQAPKLPVHVGKRAMIAADL 120  
 QY 122 ETDVDFINVKATTTKLGFBGRKEGIAVQAVLI 155  
 DB 121 GCHMDVNVKATTTKLGFBGRKEGIAVQAVLI 154  
 RESULT 7  
 ABP70685  
 ID ABP70685 standard; Protein; 159 AA.  
 AC ABP70685;  
 XX  
 DT 04-APR-2003 (first entry)  
 XX  
 DE *Escherichia coli* 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.  
 XX  
 KM Protein co-ordinate data; herbicide; protozoacide; antibacterial;  
 KM antibiotic; isoprenoid synthesis; crystal; enzyme;  
 KM 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; malaria;  
 KM Plasmodium falciparum; ispf.  
 XX  
 OS *Escherichia coli*.  
 XX  
 PN WO200292800-A2.  
 XX  
 PD 21-NOV-2002.  
 XX  
 PF 13-MAY-2002; 2002WO-BP05238.  
 XX  
 PR 15-MAY-2001; 2001DE-1023597.  
 PR 25-MAY-2001; 2001US-293875P.  
 XX  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PA (BACH/) BACHER A.  
 XX  
 PI Bacher A, Hecht S, Huber R, Kaiser J, Rohdich F, Steinbacher S;  
 XX MPI; 2003-167170/16.  
 DR N-PSDB; AAZ25497.  
 XX  
 PT Novel crystal useful for identifying a potential inhibitor of  
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, comprises



CC particular, they can be used against *M. tuberculosis*. The present  
 CC sequence is one of the proteins of the invention.

XX Sequence 158 AA;

Query Match 66.3%; Score 534; DB 21; Length 158;  
 Best Local Similarity 65.6%; Pred. No. 3.1e-55;

Matches 103; Conservative 18; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIVGCVYVHRNDGHHLLGVKIPYKGLFAHSDGVVLAALADALGAALADIGK 60  
 DB 1 MFRIGGDFVHGVDSRPLIIGIEIPYKGLFHSDDAVLAVTADACLVAGSGDICK 60

QY 61 HEPDTPNFKGADSRVLAHVGIVYKKGKLVNADVTIIAQPCKLPVPGKRIANIAD 120  
 DB 61 HEPDTPNFKGADSRVLAHVGIVYKKGKLVNADVTIIAQPCKLPVPGKRIANIAD 120

QY 121 LETDVPFINVKATTTEKLGFEGRKEGIAVAVLI 157  
 DB 121 LEADVGVVVKATTEKLGFTGRAGIAVAVLI 157

RESULT 10  
 ABP70686  
 ID ABP70686 standard; Protein; 159 AA.

XX ABP70686;  
 AC  
 XX  
 DT 04-APR-2003 (first entry)

DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase mutant DBS.

XX Protein co-ordinate data; herbicide; protozoacide; antibacterial;  
 KM antibiotic; isoprenoid synthesis; crystal; enzyme;

KM 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; malaria;

XX Plasmodium falciparum; ispf; mutant; mutein.  
 OS  
 XX Escherichia coli.  
 XX Synthetic.

XX Key Location/Qualifiers  
 FT Misc-difference 8 /note= "Wild-type Asp substituted by Ser"  
 FT

XX WO200292800-A2.

XX 21-NOV-2002.

XX 13-MAY-2002; 2002WO-EP05238.

XX 15-MAY-2001; 2001DE-1023597.

XX 25-MAY-2001; 2001US-293875P.

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX (BACH/) BACHER A.

XX Bacher A, Hecht S, Huber R, Kaiser J, Rohdich F, Steinbacher S;

XX WPI; 2003-167170/16.

XX Novel crystal useful for identifying a potential inhibitor of  
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, comprises  
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein with or  
 PT without zinc

XX Example 2; Page -, 214pp; English.

XX The present invention relates to a crystal comprising  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase with or without  
 CC zinc. The crystal is useful in an inhibitor-screening assay, and  
 CC identified inhibitors are useful as antibiotics against bacteria or  
 CC protozoa especially malarial parasite e.g. Plasmodium falciparum, and  
 CC as herbicides. The inhibitors are also useful for inhibiting the

CC synthesis of isoprenoids. The present sequence is a mutant of  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase which was generated  
 CC in an example from the invention.

CC Note: The present sequence is not shown in the specification, but is  
 CC derived from the wild-type sequence of Escherichia coli

CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ABP70685).

XX Sequence 159 AA;

Query Match 65.9%; Score 511; DB 24; Length 159;  
 Best Local Similarity 68.8%; Pred. No. 7.2e-55;  
 Matches 106; Conservative 14; Mismatches 34; Indels 0; Gaps 0;

QY 2 IRVGVGVYVHRNDGHHLLGVKIPYKGLFAHSDGVVLAALADALGAALADIGK 61  
 DB 1 MFRIGGDFVHGVDSRPLIIGIEIPYKGLFHSDDAVLAVTADACLVAGSGDICK 60

QY 62 PPDTPNFKGADSRVLAHVGIVYKKGKLVNADVTIIAQPCKLPVPGKRIANIAD 121  
 DB 61 PPDTPNFKGADSRVLAHVGIVYKKGKLVNADVTIIAQPCKLPVPGKRIANIAD 120

QY 122 ETDVPFINVKATTTEKLGFEGRKEGIAVAVLI 155  
 DB 121 GCHMDVAVVKATTEKLGFTGRAGIAVAVLI 154

RESULT 11  
 ABP70687  
 ID ABP70687 standard; Protein; 159 AA.

XX ABP70687;  
 AC  
 XX  
 DT 04-APR-2003 (first entry)

DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase mutant H108.

XX Protein co-ordinate data; herbicide; protozoacide; antibacterial;  
 KM antibiotic; isoprenoid synthesis; crystal; enzyme;

KM 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; malaria;

KM Plasmodium falciparum; ispf; mutant; mutein.

XX Escherichia coli.  
 XX Synthetic.

XX Key Location/Qualifiers  
 FT Misc-difference 10 /note= "Wild-type His substituted by Ser"  
 FT

XX WO200292800-A2.

XX 21-NOV-2002.

XX 13-MAY-2002; 2002WO-EP05238.

XX 15-MAY-2001; 2001DE-1023597.

XX 25-MAY-2001; 2001US-293875P.

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX (BACH/) BACHER A.

XX Bacher A, Hecht S, Huber R, Kaiser J, Rohdich F, Steinbacher S;

XX WPI; 2003-167170/16.

XX Novel crystal useful for identifying a potential inhibitor of  
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, comprises  
 PT 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein with or  
 PT without zinc

XX Example 3; Page -, 214pp; English.

XX The present invention relates to a crystal comprising  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase with or without

CC zinc. The crystal is useful in an inhibitor-screening assay, and  
 CC identified inhibitors are useful as antibiotics against bacteria or  
 CC protozoa especially malarial parasite e.g. Plasmodium falciparum, and  
 CC as herbicides. The inhibitors are also useful for inhibiting the  
 CC synthesis of isoprenoids. The present sequence is a mutant of  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase which was generated  
 CC in an example from the invention.  
 CC Note: The present sequence is not shown in the specification, but is  
 CC derived from the wild-type sequence of Escherichia coli  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ABP70685).

XX Sequence 159 AA;

Query Match 65.5%; Score 528; DB 24; Length 159;  
 Best Local Similarity 68.8%; Pred. No. 1.6e-54;  
 Matches 106; Conservative 14; Mismatches 34; Indels 0; Gaps 0;

QY 2 IRVAGYDVHRRFNDGHHILGGVKIPYEKGLAESDGDVVLHALLADALIGAAALGDIGKH 61  
 DB 1 KRIGHGFVSARFGEGPILIGGVRIPEYKGLAHSDDVALLTDALIGAAALGDIGKL 60  
 QY 62 FPDTPNFKGADSRVLLRHVYGIYKKGKYLVNADVTTIIAQAAPKMLPHVPGMRANIAADL 121  
 DB 61 FPDTPNFKGADSRVLLRHVYGIYKKGKYLVNADVTTIIAQAAPKMLPHVPGMRANIAADL 120  
 QY 122 ETDVDFINVKATTTTEKLGFEGRKEGIAVQAVYLI 155  
 DB 121 GCHMDVNVKATTTTEKLGFEGRKEGIAVQAVYLI 154

RESULT 12

ABP70688 standard; Protein; 159 AA.

AC ABP70688;

DT 04-APR-2003 (first entry)

DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase mutant H425.

XX Protein co-ordinate data; herbicide; protozoacide; antibacterial;

KM antibiotic; isoprenoid synthesis; crystal; enzyme;

KM 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; malarial;

KM Plasmodium falciparum; ispf; mutant; mutain.

OS Escherichia coli.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 42 /note= "Wild-type His substituted by Ser"

XX WO200292800-A2.

XX 21-NOV-2002.

XX 13-MAY-2002; 2002WO-EP05238.

XX 15-MAY-2001; 2001DE-1023597.

XX 25-MAY-2001; 2001US-293875P.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX (BACH/) BACHER A.

XX Bacher A, Hecht S, Huber R, Kaiser J, Rohdich F, Steinbacher S;

XX WPI; 2003-167170/16.

XX Novel crystal useful for identifying a potential inhibitor of

XX 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, comprises

XX 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein with or

XX without zinc

PS Example 4; Page -; 214pp; English.

XX The present invention relates to a crystal comprising  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase with or without  
 CC zinc. The crystal is useful in an inhibitor-screening assay, and  
 CC identified inhibitors are useful as antibiotics against bacteria or  
 CC protozoa especially malarial parasite e.g. Plasmodium falciparum, and  
 CC as herbicides. The inhibitors are also useful for inhibiting the  
 CC synthesis of isoprenoids. The present sequence is a mutant of  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase which was generated  
 CC in an example from the invention.  
 CC Note: The present sequence is not shown in the specification, but is  
 CC derived from the wild-type sequence of Escherichia coli  
 CC 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ABP70685).

XX Sequence 159 AA;

Query Match 65.5%; Score 528; DB 24; Length 159;  
 Best Local Similarity 68.8%; Pred. No. 1.6e-54;  
 Matches 106; Conservative 14; Mismatches 34; Indels 0; Gaps 0;

QY 2 IRVAGYDVHRRFNDGHHILGGVKIPYEKGLAESDGDVVLHALLADALIGAAALGDIGKH 61  
 DB 1 KRIGHGFVSARFGEGPILIGGVRIPEYKGLAHSDDVALLTDALIGAAALGDIGKL 60  
 QY 62 FPDTPNFKGADSRVLLRHVYGIYKKGKYLVNADVTTIIAQAAPKMLPHVPGMRANIAADL 121  
 DB 61 FPDTPNFKGADSRVLLRHVYGIYKKGKYLVNADVTTIIAQAAPKMLPHVPGMRANIAADL 120  
 QY 122 ETDVDFINVKATTTTEKLGFEGRKEGIAVQAVYLI 155  
 DB 121 GCHMDVNVKATTTTEKLGFEGRKEGIAVQAVYLI 154

RESULT 13

ABG75686 standard; Protein; 170 AA.

AC ABG75686;

DT 02-JUN-2003 (first entry)

DE 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECPs).

XX 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; MECPs; enzyme;

KM 3D molecular structure; protein co-ordinate data; antimicrobial.

XX Haemophilus influenzae.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "Featured in specification as Mse"

XX Misc-difference 69 /note= "Featured in specification as Mse"

XX Misc-difference 108 /note= "Featured in specification as Mse"

XX Misc-difference 115 /note= "Featured in specification as Mse"

XX WO2002102991-A2.

XX 27-DEC-2002.

XX 17-JUN-2002; 2002WO-US19451.

XX 13-JUN-2001; 2001US-299058P.

XX (STRU-) STRUCTURAL GENOMIX INC.

XX Louis GV, Buchanan SG, Gajiwala KS, Sauder MJ;

XX WPI; 2003-247867/24.

PT Novel 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein or  
 PT its functional protein subunit, in crystalline form, useful for  
 PT identifying and designing inhibitors and activators of the protein  
 XX  
 XX  
 PS Disclosure; Fig 2; 372pp; English.

CC The invention relates to a 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC synthase (MECPs) protein, or a functional MECPs protein subunit, in  
 CC crystalline form. The invention also relates to a computer-readable  
 CC database comprising representation of a compound capable of binding a  
 CC binding pocket of an MECPs protein. The methods are useful for producing  
 CC a compound comprising a 3D molecular structure represented by the  
 CC coordinates contained in a computer readable database, modulating MECPs  
 CC protein activity by contacting the MECPs with a compound, identifying an  
 CC activator or inhibitor of a protein that comprises an MECPs active site  
 CC or binding pocket, producing a mutant MECPs protein having an altered  
 CC property relative to an MECPs protein and determining whether a compound  
 CC binds the MECPs protein. The protein is useful for identifying and  
 CC designing inhibitors and activators of MECPs, for designing  
 CC anti-microbials that target the active site or a binding format of MECPs  
 CC or otherwise interfere with MECPs activity or another activity in an  
 CC associated biochemical, metabolic or anabolic pathway, or for rational  
 CC drug design to identify and/or design compounds that bind MECPs for  
 CC developing new therapeutic agents. This sequence represents the MECPs  
 CC protein of the invention.

XX Sequence 170 AA;

Query Match 64.4%; Score 519; DB 24; Length 170;  
 Best Local Similarity 63.7%; Pred. No. 2.1e-53;  
 Matches 100; Conservative 21; Mismatches 36; Indels 0; Gaps 0;

QY 1 MIRVGWGYDHRFNDGHIILGCVKIPYEKGLFAHSDGDVLTALADALIGAALGDIGK 60  
 DB 3 LRIGGFVQVHAFGEGRPLIGVEVPHRTGFIAHSDGVLTALADALIGAALGDIGK 62  
 QY 61 HRPDTPNFKGADSRVLLRHVYGIKVKGYKLVNADVTIIAQAPKLPVPGKRNIAADL 120  
 DB 63 LRPDTPNFKGADSRVLLRHVYGIKVKGYKLVNADVTIIAQAPKLPVPGKRNIAADL 122  
 QY 121 LETVDVFINVKATTTTEKLGPEGRKEGIAVOAVVLIIR 157  
 DB 123 LQCDIEGVVKKATTTTEKLGFTGRGEGIAVAVVLIIR 159

RESULT 14

ABP79612  
 ID ABP79612 standard; Protein; 174 AA.

XX  
 AC ABP79612;

XX  
 DT 07-MAR-2003 (first entry)

XX  
 DE N. gonorrhoeae amino acid sequence SEQ ID 5754.

XX  
 KW Antibacterial; infection; vaccine; gene therapy.

XX  
 OS Neisseria gonorrhoeae.

XX  
 PN WO200279243-A2.

XX  
 PD 10-OCT-2002.

XX  
 PF 12-FEB-2002; 2002WO-1B02069.

XX  
 PR 12-FEB-2001; 2001GB-0003424.

XX  
 PA (CHIR-) CHIRON SPA.

XX  
 PI Fontana MR, Pizza M, Massignani V, Monaci E;

XX  
 DR WPI, 2003-058415/05.

XX  
 DR N-PSDB; AB240582.

XX  
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection  
 XX  
 XX  
 PS Disclosure; Page 608; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention.

XX Sequence 174 AA;

Query Match 62.9%; Score 507; DB 24; Length 174;  
 Best Local Similarity 60.9%; Pred. No. 6e-52;  
 Matches 95; Conservative 22; Mismatches 39; Indels 0; Gaps 0;

QY 2 IRVGWGYDHRFNDGHIILGCVKIPYEKGLFAHSDGDVLTALADALIGAALGDIGK 61  
 DB 18 LRIGGFVQVHAFGEGRPLIGVEVPHRTGFIAHSDGVLTALADALIGAALGDIGK 77  
 QY 62 FPDTPNFKGADSRVLLRHVYGIKVKGYKLVNADVTIIAQAPKLPVPGKRNIAADL 121  
 DB 78 FPDTPNFKGADSRVLLRHVYGIKVKGYKLVNADVTIIAQAPKLPVPGKRNIAADL 137  
 QY 122 ETDVDVFINVKATTTTEKLGPEGRKEGIAVOAVVLIIR 157  
 DB 138 GDISCVNIKKTNEKLGFTGRGEGIAVAVVLIIR 173

RESULT 15

ABU06061  
 ID ABU06061 standard; Protein; 160 AA.

XX  
 AC ABU06061;

XX  
 DT 22-JAN-2003 (first entry)

XX  
 DE N. meningitidis vaccine antigen #40.

XX  
 KW Vaccine; antigen; meningococcal disease; pathogenic bacteria;

XX  
 OS Neisseria meningitidis group B.

XX  
 PN WO200277648-A2.

XX  
 PD 03-OCT-2002.

XX  
 PF 22-MAR-2002; 2002WO-GB01399.

XX  
 PR 22-MAR-2001; 2001GB-0007219.

XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX  
 PI (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX  
 PI Robinson AJ, Gorrings AR, Hudson MJ, Bracegirdle P, West DM;

XX  
 DR Oliver KJ, Kroll JS, Langford PR;

XX  
 DR WPI, 2003-018958/01.

XX  
 DR N-PSDB; ABX09898.

XX  
 PT Identifying an antigen for manufacturing a vaccine against  
 PT meningococcal infection, comprises contacting antibodies with  
 PT polypeptides, detecting polypeptide-antibody complexes, and identifying  
 PT bound polypeptides as antigens -

XX  
 PS Claim 37; Page 237-238; 310pp; English.

CC The invention relates to identifying an antigen comprising:

CC (a) obtaining antibodies against a commensal bacteria; or an extract

CC from a commensal bacteria; (b) contacting the antibodies with

CC polypeptides obtained from an expression library of either a commensal or

CC a pathogenic bacteria; (c) determining whether the polypeptides bind to

CC antibodies; and (d) (where a polypeptide binds to an antibody)

CC identifying that polypeptide as an antigen. Also included are:

CC (1) a method of preparing a vaccine composition, comprising identifying

CC an antigen with the above method, and combining the antigen with a

CC carrier; (2) a vaccine composition obtained by the above methods;

CC (3) an antigen identified by the above methods; (4) a polypeptide

CC encoded by all or a part of a nucleic acid sequence comprising the

CC Neisseria lactamica DNA sequences detailed in the specification;

CC (5) a vector comprising the nucleic acid molecule; (6) a method of

CC preparing a composition for vaccination against infection by pathogenic

CC bacteria, comprising: (a) obtaining a first antigen from a commensal

CC Neisseria; (b) comparing the amino acid sequence of the first antigen

CC with the amino acid sequence of the second antigen from a pathogenic

CC bacteria, or comparing the sequence of the nucleic acid which codes for the

CC first antigen with the sequence of the nucleic acid that codes for the

CC second antigen; and if the first antigen is homologous to the second

CC antigen or if the nucleic acid sequence for the first antigen is

CC homologous to that of the second antigen, and (c) preparing a composition

CC for vaccination against bacterial infection comprising the first antigen;

CC and (7) an antibody that binds to the polypeptide antigen.

CC The method is useful in screening commensal and pathogenic

CC bacteria for previously unidentified vaccine antigens by identifying

CC polypeptide antigens that bind to sera raised against commensal

CC bacterial proteins. The polypeptide is useful as a vaccine antigen

CC which may be used in the manufacture of a medicament for vaccination

CC against meningococcal infection (e.g. meningitis). The present

CC sequence represents an antigenic protein from the pathogenic bacteria N.

CC meningitidis.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 160 AA:

Query Match 62.7%; Score 505; DB 24; Length 160;

Best Local Similarity 60.9%; Pred. No. 9.1e-52;

Matches 95; Conservative 21; Mismatches 40; Indels 0; Gaps 0;

QY 2 IRVKGIVYHRPNDDHIIIGGVKIPYKGLFANSDGVVTLNADALGAAAGDIGN 61  
DB 4 IRIGGYVYVHQLTGRKLLIGVEIPFEGKLGSHSDADALAHAVTDALLGAAGLDSGH 63  
QY 62 FPDTPNFKGAADSVLRHVVYIVVEKGYKLVNADVTIIAQPKMLPHVPGKRAIADL 121  
DB 64 FPDIAAFKDAADSVLRHVVYIVVEKGYKLVNADVTIIAQPKMLPHVPGKRAIADL 123  
QY 122 ETDVDFINVKATTTKLGFBGRKGGIIVQAVVLIIR 157  
DB 124 GIDISCVNIKGTNKKLGYLGRMEGIEAQAVALIVR 159

Search completed: January 29, 2004, 15:49:43  
Job time : 15.3511 secs